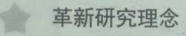


格致方法·定量研究系列 吴晓刚 主编

贝叶斯统计推断

[美] 古特蒙德·R.艾弗森（Gudmund R.Iversen）著
贺光烨 译 范新光 闵尊涛 张柏杨 于皓 校



革新研究理念

丰富研究工具

最权威、最前沿的定量研究方法指南

69

格致出版社 上海人民出版社

格致方法 · 定量研究系列　吴晓刚　主编

贝叶斯统计推断

[美] 古德蒙德 · R.艾弗森(Gudmund R. Iversen) 著
贺光烨 译　范新光　闵尊涛　张柏杨　于皓 校

SAGE Publications, Inc.

格致出版社　上海人民出版社

图书在版编目(CIP)数据

贝叶斯统计推断/(美)古德蒙德·R.艾弗森著；
贺光烨译.—上海：格致出版社：上海人民出版社，
2019.11
(格致方法.定量研究系列)
ISBN 978-7-5432-2876-4

Ⅰ.①贝… Ⅱ.①古… ②贺… Ⅲ.①贝叶斯推断-
统计推断-研究 Ⅳ.①O212

中国版本图书馆 CIP 数据核字(2018)第 112641 号

责任编辑 贺俊逸

格致方法·定量研究系列
贝叶斯统计推断
[美]古德蒙德·R.艾弗森 著
贺光烨 译
范新光 闵尊涛 张柏杨 于 皓 校

出 版	格致出版社 上海人民出版社 (200001 上海福建中路 193 号)
发 行	上海人民出版社发行中心
印 刷	浙江临安曙光印务有限公司
开 本	920×1168 1/32
印 张	4.25
字 数	75,000
版 次	2019 年 11 月第 1 版
印 次	2019 年 11 月第 1 次印刷

ISBN 978-7-5432-2876-4/C·203
定 价 32.00 元

出版说明

由香港科技大学社会科学部吴晓刚教授主编的“格致方法·定量研究系列”丛书，精选了世界著名的 SAGE 出版社定量社会科学研究丛书，翻译成中文，起初集结成八册，于 2011 年出版。这套丛书自出版以来，受到广大读者特别是年轻一代社会科学工作者的热烈欢迎。为了给广大读者提供更多的方便和选择，该丛书经过修订和校正，于 2012 年以单行本的形式再次出版发行，共 37 本。我们衷心感谢广大读者的支持和建议。

随着与 SAGE 出版社合作的进一步深化，我们又从丛书中精选了三十多个品种，译成中文，以飨读者。丛书新增品种涵盖了更多的定量研究方法。我们希望本丛书单行本的继续出版能为推动国内社会科学定量研究的教学和研究作出一点贡献。

总 序

2003年，我赴港工作，在香港科技大学社会科学部教授研究生的两门核心定量方法课程。香港科技大学社会科学部自创建以来，非常重视社会科学研究方法论的训练。我开设的第一门课“社会科学里的统计学”(Statistics for Social Science)为所有研究型硕士生和博士生的必修课，而第二门课“社会科学中的定量分析”为博士生的必修课(事实上，大部分硕士生在修完第一门课后都会继续选修第二门课)。我在讲授这两门课的时候，根据社会科学研究生的数理基础比较薄弱的特点，尽量避免复杂的数学公式推导，而用具体的例子，结合语言和图形，帮助学生理解统计的基本概念和模型。课程的重点放在如何应用定量分析模型研究社会实际问题上，即社会研究者主要为定量统计方法的“消费者”而非“生产者”。作为“消费者”，学完这些课程后，我们一方面能够读懂、欣赏和评价别人在同行评议的刊物上发表的定量研究的文章；另一方面，也能在自己的研究中运用这些成熟的方法论技术。

上述两门课的内容，尽管在线性回归模型的内容上有少

量重复，但各有侧重。“社会科学里的统计学”从介绍最基本的社会研究方法论和统计学原理开始，到多元线性回归模型结束，内容涵盖了描述性统计的基本方法、统计推论的原理、假设检验、列联表分析、方差和协方差分析、简单线性回归模型、多元线性回归模型，以及线性回归模型的假设和模型诊断。“社会科学中的定量分析”则介绍在经典线性回归模型的假设不成立的情况下的一些模型和方法，将重点放在因变量为定类数据的分析模型上，包括两分类的 logistic 回归模型、多分类 logistic 回归模型、定序 logistic 回归模型、条件 logistic 回归模型、多维列联表的对数线性和对数乘积模型、有关删节数据的模型、纵贯数据的分析模型，包括追踪研究和事件史的分析方法。这些模型在社会科学研究中有着更加广泛的应用。

修读过这些课程的香港科技大学的研究生，一直鼓励和支持我将两门课的讲稿结集出版，并帮助我将原来的英文课程讲稿译成了中文。但是，由于种种原因，这两本书拖了多年还没有完成。世界著名的出版社 SAGE 的“定量社会科学研究”丛书闻名遐迩，每本书都写得通俗易懂，与我的教学理念是相通的。当格致出版社向我提出从这套丛书中精选一批翻译，以飨中文读者时，我非常支持这个想法，因为这从某种程度上弥补了我的教科书未能出版的遗憾。

翻译是一件吃力不讨好的事。不但要有对中英文两种语言的精准把握能力，还要有对实质内容有较深的理解能力，而这套丛书涵盖的又恰恰是社会科学中技术性非常强的内容，只有语言能力是远远不能胜任的。在短短的一年时间里，我们组织了来自中国内地及香港、台湾地区的二十几位

研究生参与了这项工程，他们当时大部分是香港科技大学的硕士和博士研究生，受过严格的社会科学统计方法的训练，也有来自美国等地对定量研究感兴趣的博士研究生。他们是香港科技大学社会科学部博士研究生蒋勤、李骏、盛智明、叶华、张卓妮、郑冰岛，硕士研究生贺光烨、李兰、林毓玲、肖东亮、辛济云、於嘉、余珊珊，应用社会经济研究中心研究员李俊秀；香港大学教育学院博士研究生洪岩璧；北京大学社会学系博士研究生李丁、赵亮员；中国人民大学人口学系讲师巫锡炜；中国台湾"中央"研究院社会学所助理研究员林宗弘；南京师范大学心理学系副教授陈陈；美国北卡罗来纳大学教堂山分校社会学系博士候选人姜念涛；美国加州大学洛杉矶分校社会学系博士研究生宋曦；哈佛大学社会学系博士研究生郭茂灿和周韵。

参与这项工作的许多译者目前都已经毕业，大多成为中国内地以及香港、台湾等地区高校和研究机构定量社会科学方法教学和研究的骨干。不少译者反映，翻译工作本身也是他们学习相关定量方法的有效途径。鉴于此，当格致出版社和 SAGE 出版社决定在"格致方法 · 定量研究系列"丛书中推出另外一批新品种时，香港科技大学社会科学部的研究生仍然是主要力量。特别值得一提的是，香港科技大学应用社会经济研究中心与上海大学社会学院自 2012 年夏季开始，在上海（夏季）和广州南沙（冬季）联合举办《应用社会科学研究方法研修班》，至今已经成功举办三届。研修课程设计体现"化整为零、循序渐进、中文教学、学以致用"的方针，吸引了一大批有志于从事定量社会科学研究的博士生和青年学者。他们中的不少人也参与了翻译和校对的工作。他们在

繁忙的学习和研究之余，历经近两年的时间，完成了三十多本新书的翻译任务，使得“格致方法·定量研究系列”丛书更加丰富和完善。他们是：东南大学社会学系副教授洪岩璧，香港科技大学社会科学部博士研究生贺光烨、李忠路、王佳、王彦蓉、许多多，硕士研究生范新光、缪佳、武玲蔚、臧晓露、曾东林，原硕士研究生李兰，密歇根大学社会学系博士研究生王骁，纽约大学社会学系博士研究生温芳琪，牛津大学社会学系研究生周穆之，上海大学社会学院博士研究生陈伟等。

陈伟、范新光、贺光烨、洪岩璧、李忠路、缪佳、王佳、武玲蔚、许多多、曾东林、周穆之，以及香港科技大学社会科学部硕士研究生陈佳莹，上海大学社会学院硕士研究生梁海祥还协助主编做了大量的审校工作。格致出版社编辑高璇不遗余力地推动本丛书的继续出版，并且在这个过程中表现出极大的耐心和高度的专业精神。对他们付出的劳动，我在此致以诚挚的谢意。当然，每本书因本身内容和译者的行文风格有所差异，校对未免挂一漏万，术语的标准译法方面还有很大的改进空间。我们欢迎广大读者提出建设性的批评和建议，以便再版时修订。

我们希望本丛书的持续出版，能为进一步提升国内社会科学定量教学和研究水平作出一点贡献。

吴晓刚

于香港九龙清水湾

序

早在该系列丛书出版以前,我们就基于经典推理统计发表了一篇关于显著性检验的论文。在本书中,艾弗森教授就贝叶斯推断表达了另一种观点,即,基于先验概率生成后验概率。前四章,艾弗森教授以温和清晰的笔调介绍了贝叶斯方法,同时运用了一些简单范例来帮助初学者理解该方法。

在阅读本书之前,读者应该对经典统计推论,尤其是对基础概率论和二项分布有一定了解。初学者要做好相应准备,从概念和计算上仔细研读和思考第 5 章里的有关材料和举出的例子。如果同学们愿意投资这个时间成本,接下来的第 6 章至第 8 章的学习就会非常有意义。当整本书被消化过后,就可以算是入门了。

在本书中,艾弗森教授讲解了如何运用贝叶斯理论及统计推论估计各种参数(包括比例、均值、相关性、回归和方差)。且在每个例子中,他都对贝叶斯推论法和经典推

论法进行了比较。在后面的章节，他还讲述了贝叶斯推论的相关优缺点，最终得出结论：贝叶斯方法优于经典推论法。不论对此结论同意与否，通过认真学习，读者们均会对贝叶斯统计方法所涉及的内容及其优缺点有一个更深的了解。

在这本书里，艾弗森教授就贝叶斯统计的一个主要问题——先验概率的主观性进行了批判，该问题的重要性可与经典统计中假设检验的置信区间（confidence interval）选择相提并论。另外，他对先验分布的相对影响的处理，以及在实际数据中获取有关后验概率的信息也进行了详述。同时，艾弗森教授还检验了不同条件下这两类要素的敏感性，有助于新学者理解，相比实地搜集来的分析数据，根据先验信息作出的主观假设的相对效用如何。值得提及的是，就置信区间而言，事实上，经典统计方法与假设均匀先验分布的贝叶斯方法一样，同样可以得出一致的数字结果。然而，两者就结果的解释有所不同，因此研究者可能要用常规方法计算结果，而如果该先验分布是有意义的，则用贝叶斯方法进行解释。通读这本书后，读者可能会发觉贝叶斯方法的独特魅力及其在社会科学中强大的应用潜力。

约翰·L.苏利文

目　录

第 1 章

托马斯·贝叶斯统计推断*

贝叶斯统计提供了一种假设性检验及置信区间估计的替换方法。该方法是以英国牧师托马斯·贝叶斯(Thomas Bayes，1760年去世)命名的。贝叶斯于1763年发表的一篇关于概率等价性的论文就是现今广为所知的贝叶斯定理。当论文第一次发表时，没有人知道这个简单相等除了在一些与概率论相关的问题上运用外还可以在哪里运用。然而，两百年后贝叶斯统计被赋予了新的意义，并奠定了贝叶斯统计推论的基础。

统计推论是指将从已知样本中获得的结论推断到未知总体中去的一种方法。例如，我们知道在样本中有55%的选举人以某种方式进行投票，那么在总体中究竟有多少选举人会以这种方式投票呢？要对总体百分比进行推断，

---

* 本书得益于鲍比·艾弗森(Bobbie Iversen)和两位审稿人的意见以及娜奥米·马尔库斯(Naomi Marcus)的审阅。作者感谢拉里·艾默尔(Larry Ehmer)提供第4章的图表，以及政治和社会研究跨校夏令营提供的第5章的数据。美国生活质量的调查数据(1978)是由密歇根大学社会研究所政治研究中心所搜集的，受到了国家科学基金的资助。无论是数据搜集者还是夏令营都不承担本书分析或解释的责任。

一种方法是就总体百分比设定一个零假设，如，总体百分比为50%，然后用样本数据来检验该假设是否可以被拒绝。一个常用步骤就是通过样本数据对未知总体百分比估计一个置信区间。或者，我们也可以通过贝叶斯统计来推断总体百分比。

统计推断不论采取何种方法均需要包括一个通过已知样本对更大的未知总体推断的过程。因为是推断，所以我们永远无法肯定所得出结论是百分百正确的。该不确定性就需通过概率来处理，这些概率可以被称为显著性水平和置信水平，也可以是贝叶斯先验概率和贝叶斯后验概率。

选择哪种统计推论主要基于对概率这个概念的定义。概率的数学理论并不关心概率本身是如何被量度的，但是若要用概率，我们就需要按照一定方法对其进行赋值。如何量度概率决定了在进行统计推论时，应该用假设检验、置信区间估计还是贝叶斯统计。这两种相关的概率可以被分别称为经验概率和主观概率。还存在第三种概率，即必要或逻辑概率，这里不予讨论，感兴趣的读者请参见巴内特(Barnett，1982)。

从经验性的角度来看，概率是一种长期的相对频率或比例，因此常被称为频率概率或者客观概率。假设对一个实验重复多次，每次都会发生一个事件。当实验次数越来越多，某个事件出现次数的比例就趋近于该事件的发生概

率。以掷骰子为例,若掷的次数足够多,出现1或2的次数比例大约为所有投掷次数的1/3,因此可以得出结论,出现1或2的概率为1/3。该观点第一次被韦恩(Venn, 1886)提出,之后便在经典统计推断的纽曼-皮尔森(Neyman-Pearson)系统里进行假设检验和置信区间估计时得以应用。

统计零假设检验的显著性水平就是一个相对频率的例子。如果零假设为真且在同一总体中抽取了许多不同的样本,那么总有一些样本会落入检验的拒绝区间内。这些样本就会使我们错误得拒绝零假设。其中,落入拒绝区间次数的长期比例就是检验的显著性水平。有关此类统计推论相关问题的批判性检验,请见第2章。

从主观角度来看,概率是个人基于已有证据对不确定性的量度。由于大多证据是经验性的,因此,不论是从经验角度还是主观角度,对于某一个特定概率,所得概率数值均相同。对于前例,若从主观角度来看,(基于掷骰子的物理特性)一个公平骰子出现1或2的概率也为1/3。因掷骰子所出现事件的概率密度函数服从均匀分布,据经验发现,掷骰子得到1或2的比例是1/3,所以,对于一个公平骰子,其出现1或2的概率就为1/3。

我们还可以将主观概率运用到无法重复的事件上。政治学家常常基于当前的政治形势,例如,该总统对再次参选的态度等,来表达第一个任期的总统再次参选的概

率。我们无法一直观察该总统的行为，将时间调回并记录下他每一次的决定。因此，对于该总统是否参选的概率则没有经验性和实验性基础。任何使用经典统计推论方法的人都无法用概率来考虑第二次竞选的可能性。由于主观概率可以用于量度单个唯一事件的不确定性，因此主观概率相较经验概率更具有一般性，且这些概率并不唯一。由于两个理性的人对于总统意图的判断基于的信息可能会不同，因此有关总统是否寻求第二任期的主观概率可能不同。

主观概率的开创性研究始于拉姆瑟(Ramsey, 1926)，20 世纪 30 年代蒂菲尼提(deFinetti)以及萨瓦吉(Savage, 1954)也做了大量工作。有关概率的一些主观见解的讨论，除了蒂菲尼提著名的论文之外，还可参考凯伯格和斯莫克勒(Kyburg & Smokler, 1980)，蒂菲尼提(deFinetti, 1982)，以及杰弗里(Jeffrey, 1983)的研究。

第2章

经典统计推断

现今社会科学中大多统计推断均是假设检验某一形式的运用，相关论述请参见亨克尔(Henkel，1976)。尽管这种方法相当流行，我们不得不承认经典统计推断着实存在一些缺陷。这些缺陷可以通过贝叶斯方法在一定程度上得到弥补。为了激发读者兴趣并令贝叶斯统计更易理解，首先我们要对经典方法的相关特征进行批判性审视。更多有关贝叶斯理论和贝叶斯统计的介绍会在第 3 章涉及。读者们若想了解有关更多不同统计推断的方法，请参见巴内特(Barnett，1982)。

第1节 | 运用尾概率

假设我们想要研究某总体是否在性别间均匀地分布。那么，该研究问题的统计零假设则为在随机抽取的人中性别为女的概率为0.5。基于该零假设，我们抽取一个样本量为10的随机样本，并做双尾检验(two-sided test)。若零假设为真，那么通过二项分布(binomial distribution)，可知道样本中包含0个女性(和10个男性)的概率为1/1 024。同样，样本中包含1个女性(和9个男性)的概率为10/1 024，1个男性(和9个女性)也为10/1 024。因此，若我们拒绝包含0、1、9，或者10个女性的假设，那么检验的显著性水平等于相应概率的总和，即，22/1 024=0.02。

现在我们假设所用样本的样本量为10，且其中有9个女性。若所得结果落入检验的拒绝区间，我们即可得出结论：在0.02显著性水平下，零假设被拒绝。该显著性水平包含部分观测数据的概率，似乎将样本数据作为是否拒绝零假设的决定所依赖证据的一部分也非常合理。然而值得注意的是，该显著性水平还包括含有0、1，或10个女性

的概率，但这些情况并未发生。由于数据中那些并未发生的事件的概率也作为了拒绝零假设的部分证据，因此经典统计推断常令我们处于一个非常尴尬的境地。

会陷入这种困境的原因在于，经典统计推断理论是将概率作为长期的相对频率。显著性水平表明若抽取大量样本，从长远来看会发生什么情况。若零假设为真，但我们发现2%的样本落入了拒绝区间，这样就导致了零假设被错误地拒绝了。因此，在数据收集之前，显著性水平作为一种概率是一个非常有意义的特性。问题在于我们并没有很多样本，仅有一个。因此在样本已知的情况下，相对于一个样本，解释显著性水平就愈发困难。

第2节｜置信区间的解释

与在样本数据已知的情况下,对置信区间的解释也会面临与概率使用方面相同的困境。相关理论表明从长远来看,通过多样本数据可以得到大量的置信区间,但是只有一定比率的置信区间会包含参数真值而余下的不会。因此,在数据收集之前,我们可经由理论预测从长远来看未来会发生什么。然而如前所述,在已知样本的情况下,如何处理数据中所得概率就变得非常困难。通过样本所得的置信区间可以包含或不包含参数真值,但是我们不知道所得的那一个置信区间是否属于包含参数真值的区间集合中的一个,我们只是抱有希望。

第3节 参数值的不确定性

有关尾概率和置信区间的讨论均对经典统计推断提出了挑战。统计推断的目的是研究未知总体的参数值。由于对参数值不确定，为了了解这些参数我们去收集数据。然而只有样本数据，我们不能奢望找到确切的参数值，在对样本数据进行研究之后，我们或多或少可以对该参数有更多的了解。所以在样本数据收集之前，参数值的不确定性较强，然而通过从样本中不断挖掘出的新证据，便可知道更多有关参数的特性。即便这些信息仍无法令我们确定参数值，但随着对样本数据的了解程度的增加，该不确定性会一定程度地降低。因此我们需要的就是这样一种统计推断方法，始于对参数的不确定，通过收集的样本数据所含信息对该不确定性不断修改。

有关经典统计的详细讨论可以参考一些贝叶斯统计学家的著作，如伯杰(Berger, 1980)、爱德华兹等人(Edwards et al., 1963)及罗森克兰兹(Rosenkrantz, 1977)。

第3章

贝叶斯定理

第 1 节 | 推导

对某些概率之间相等关系的正式推导，即贝叶斯定理，它是基于两个事件 P 和 D 的联合概率可以写成一个事件的概率和另一个事件的条件概率的乘积的这一事实产生的。用标号表示为：

$$\mathrm{Prob}(PD)=\mathrm{Prob}(P)\cdot\mathrm{Prob}(D|P) \qquad [3.1]$$

通过颠倒两个事件，两个事件的联合概率也可以被写作：

$$\mathrm{Prob}(DP)=\mathrm{Prob}(D)\cdot\mathrm{Prob}(P|D) \qquad [3.2]$$

由于等式 3.1 和等式 3.2 等号左边相同，将式 3.1 代入式 3.2 变换后得：

$$\mathrm{Prob}(P|D)=\frac{\mathrm{Prob}(D|P)\cdot\mathrm{Prob}(P)}{\mathrm{Prob}(D)} \qquad [3.3]$$

然而能存在不止一个 P，我们将多个 P 分别命名为 P1，P2，…，Pk，且它们之间互斥且详尽。这种情况下，分母中的概率 D 可以表达为加权后的条件概率 Prob(D|Pi) 的和，权重为 Prob(Pi)。

对于事件 Pi,以上等式可写为:

$$\text{Prob}(\text{Pi}|\text{D})=\frac{\text{Prob}(\text{D}|\text{Pi})\text{Prob}(\text{Pi})}{\text{Prob}(\text{D}|\text{P1})\text{Prob}(\text{P1})+\cdots+\text{Prob}(\text{D}|\text{Pk})\text{Prob}(\text{Pk})} \quad [3.4]$$

这就是包含 k 个不同 P 的离散事件的贝叶斯定理。

字母 P 和 D 的选择是存在目的性的。其中,P 表示总体,D 表示数据。贝叶斯定理的左边为给定观测数据 D,第 i 个总体 Pi 的概率;Prob(D|Pi)表示给定总体 Pi,观测数据 D 的概率。而 Prob(Pi)为在未知数据下 Pi 的概率。对于该分析,我们所要求的是等式左边的条件概率,这一数值可以通过上式得到。

第2节 范例

假设有3个不同的社区，分别用P1、P2及P3表示3个不同的总体。这里我们想知道样本数据分别来自哪个社区。已知在P1中，30%的人为天主教徒；在P2中，相应的比例为50%，而在P3中，该比例为70%。我们通过掷骰子的形式随机选择其中一个社区，若骰子出现1或2用P1，3或4用P2，5或6用P3。在所选的社区里，我们再随机选取一个人作为一个随机样本。假设这个人是天主教徒，那么数据D包含一个天主教徒的样本。问题在于我们只知道观测数据，而并不知道所选取的是哪一个社区。因此，这个人来自P1、P2和P3这三个社区的概率是多少是这里我们所感兴趣的。

由于先前社区的选择是通过等概率随机抽取的，每个社区都有1/3的概率被选中，因此，Prob(P1)=Prob(P2)=Prob(P3)=0.333。这就是所谓先验概率，因为三个总体的概率在知道数据之前就明确了。

要用贝叶斯定理，另外我们需要知道的概率就是三个

条件概率，Prob(D|P1)、Prob(D|P2)及 Prob(D|P3)。若该人来自 P1，因为 P1 包含 30%的天主教徒，那么 Prob(D|社区 1)就应该等于 0.3。若该人来自 P2，那么 Prob(D|社区 2)为 0.5。同样，若该人来自 P3，相应的条件概率 Prob(D|社区 3)则为 0.7。因此，三个数据概率分别为：

$$\text{Prob}(D|P1)=0.3$$

$$\text{Prob}(D|P2)=0.5$$

$$\text{Prob}(D|P3)=0.7$$

基于贝叶斯定理，我们有：

$$\text{Prob}(P1|D)=\frac{0.3(0.333)}{0.3(0.333)+0.5(0.333)+0.7(0.333)}=\frac{0.0999}{0.4995}=0.20$$

$$\text{Prob}(P2|D)=\frac{0.5(0.333)}{0.3(0.333)+0.5(0.333)+0.7(0.333)}=\frac{0.1665}{0.4995}=0.33$$

$$\text{Prob}(P3|D)=\frac{0.7(0.333)}{0.3(0.333)+0.5(0.333)+0.7(0.333)}=\frac{0.2331}{0.4995}=0.47$$

这些就是后验概率。以上数值告诉我们，给定数据中的人为天主教徒，那么该人来自 P1 的数据概率为 0.20，来自 P2 的数据概率为 0.33，来自 P3 的数据概率为 0.47。容

易发现,在我们知道数据之前,来自任意一个社区的概率为0.33,而基于数据,两个概率已经有了变化。基于包括一个天主教徒的已知样本数据,该人来自P3的概率是来自P1的两倍以上,且比来自P2的概率高出约50%。

相关计算可见表3.1。第一列表明总体类别,第二列给出了总体参数,即,天主教徒的比例。第三列是各种总体参数的先验概率,且所有数值之和为1。第四列是观测数据的概率,其值取决于所用的总体是哪一个。第五列给出了数据概率与每个总体先验概率的乘积。对于不同总体,这三个乘积分别对应贝叶斯定理中的分子,而乘积之和为分母。最后一列为所得的后验概率,其可通过前一列的每个乘积除以前一列乘积之和得出。

表3.1 贝叶斯定理计算案例

总体P	天主教徒	先验概率	数据概率	乘积	后验概率
	%	Prob(Pi)	Prob(D|Pi)	Prob(D|Pi)Prob(Pi)	Prob(Pi|D)
(1)	(2)	(3)	(4)	(5)=(4)(3)	(6)
1	30	0.333	0.3	0.099 9	0.20
2	50	0.333	0.5	0.166 5	0.33
3	70	0.333	0.7	0.233 1	0.47
	总计	0.999	总计	0.499 5 =P(D)	总计 1.00

我们将对该示例的讨论推进一步,并收集更多数据。在这样一个阶段,我们关于样本来自哪个总体的不确定性可以由0.20、0.33和0.47这三个概率反映。至此,P3更可能是数据来源,但是其他两个总体的概率也不低,因此可

能不足以将其排除。从而,取得更加准确结果的唯一方法就是收集更多的信息。

新数据包含另一个样本,其也是由之前的未知总体所生成。该新样本包含10个人,其中8个为天主教徒。对于三个总体,其先验概率即之前的后验概率分别为0.20、0.33和0.47。数据概率可以通过二项分布得到。如果这些数据来自第一个总体,其中30%的人为天主教徒,那么样本量为10,且其中包含8个天主教徒和2个非天主教徒的数据概率为:

$$\text{Prob}(\text{Data}|\text{P1}) = \begin{bmatrix}10\\ \\8\end{bmatrix} 0.3^8(1-0.3)^2 = 0.001\,45 \quad [3.5]$$

同样,若数据来自第二个总体,数据概率为:

$$\text{Prob}(\text{Data}|\text{P2}) = \begin{bmatrix}10\\ \\8\end{bmatrix} 0.5^8(1-0.5)^2 = 0.043\,94 \quad [3.6]$$

若数据来自第三个总体,则

$$\text{Prob}(\text{Data}|\text{P3}) = \begin{bmatrix}10\\ \\8\end{bmatrix} 0.7^8(1-0.7)^2 = 0.233\,47 \quad [3.7]$$

将数据概率与贝叶斯定理中的先验概率合并即可生成后验概率。将数字代入后可得数据的后验概率:

$$\text{Prob}(\text{P1}|\text{Data}) = 0.002$$

$$\text{Prob}(\text{P2}|\text{Data}) = 0.115$$

$$\text{Prob}(\text{P3}|\text{Data})=0.883$$

通过这些后验概率我们对数据来源的不确定性大大减小了。很明显，数据几乎不可能来自 P1，其最可能的来源为 P3。

贝叶斯分析是累积的。如上例，最初的先验分布是基于仅含一个观测的样本。然后由该先验分布求得的后验分布又成为下一个分析（包含 10 个观测的样本）的先验分布。若仍从最初的先验分布开始，使用包含 11 个观测其中 9 个为天主教徒的样本信息，那么我们所得到最终后验分布将会相同。

第3节 | 真总体

已知存在3个不同的总体,但是只有一个总体为真。我们还知道该总体中的天主教徒百分比可能为30,50或者70。与之前相比,这种情况有何不同呢?当存在3个总体作为选择的时候,我们可通过掷骰子来决定数据来源,对于3个总体,客观存在3个有关天主教徒百分比的真实总体值,且每个总体有相同的概率被抽中。而现在只有1个总体,即,存在1个有关天主教徒百分比的真实总体值。但是问题在于不论存在3个总体还是1个,我们都无法确定百分比真值是多少。既然对于存在3个总体的情况我们都可以用贝叶斯分析,那么对于1个总体的情况,该方法也应该适用。

就两种情况而言,最主要的区别在于先验分布类型。在前一种情况下,即存在3个总体时,我们可以通过掷骰子直接确定每个总体被抽中的概率(由于被抽取概率相等,因此等于1/3),并以该概率作为数据来源的先验分布。而对于后一种情况,由于只有1个总体,则不可以通过掷骰子

来决定天主教徒百分比。然而，若概率只是不确定性的一个个人量度，我们要做的就是对这 3 个可能百分比的不确定性用 3 个概率表达出来，并令其之和为 1。做贝叶斯分析时，我们可能并不了解这个特定总体，从而有可能得出结论——3 个百分比出现的概率相等。基于此，1/3 可以进一步转化为天主教徒百分比的先验概率。

其实先验概率并不存在对与错。由于每个人所倾向的概率组合不同，选择哪种组合完全取决于我们对总体中宗教分布的认识。正因为先验概率所反映的是一种个人对不确定性的量度，所以其取值因人而异。有关不同先验分布的影响在第 6 章会有所提及。

回到刚才，我们将 3 个可能百分比的先验概率均设定为 1/3 后，余下的贝叶斯分析部分就是机械性地重复。通过数据可以发现 11 个人中有 9 个天主教徒，因而后验概率为：

$$\text{Prob}(30\% \mid \text{Data}) = 0.002$$

$$\text{Prob}(50\% \mid \text{Data}) = 0.115$$

$$\text{Prob}(70\% \mid \text{Data}) = 0.883$$

不同于前，我们处理的不是天主教徒比例不同的 3 个总体，而是每个总体都有一个自己的天主教徒百分比，因此所得到的是每个比例的后验概率，而不是来自每个总体的后验概率。

数据改变了我们对总体中天主教徒未知比例的不确定性。从1/3等概率假设开始，现在，我们相信总体中有30%的天主教徒的概率仅为0.002，含有50%天主教徒的概率稍大些为0.115，而含有70%天主教徒的概率最大，为0.883。

总体百分比正如其他总体特征（如，均值、方差以及回归系数）一样，也是一个总体参数。贝叶斯统计推断将概率分配给可能的参数值，通过贝叶斯定理，这些概率根据数据中包含的证据不断更新。在贝叶斯分析中，由于参数值是未知的，从而在一定程度上可将参数视为一个具有自身概率分布的变量。如前例，天主教徒百分比被视作含有3个可能值的离散变量。同样，参数也可被视作连续变量。对任一种情况，参数都存在一个特定的真值。当不确定该特定值的大小时，只需将参数看作一个变量并用概率来表达该参数值的不确定性。

第4章

有关比例的贝叶斯方法

有关如何用贝叶斯分析程序学习各种总体参数在这一章和下一章会有所涉及。每个分析程序都需要研究者对未知参数设定一个先验分布。那么此时我们将面临的问题是如何找到这个先验分布,以及这个分布对于分析到底有多重要。

在阐释这些问题之前,我们首先要对贝叶斯方法有一定了解,因此,在这一章,我们将首先介绍贝叶斯方法,而有关先验分布的更为具体讨论会在第 6 章介绍(如读者在接触具体方法前对先验分布感兴趣,可以先阅读第 6 章相关内容)。

第1节 | 比例

抽样调查的目的是估计总体有多大比例具有某种特征。例如,每月由政府调查所得出的当月失业率,其常常成为社会焦点;收看不同类型电视节目观众的比例将决定商业广告的费用和电台以及网络的利润。还有,总体中,对当前议题持不同看法的人口的比例决定了公众舆论,反过来也将影响政策制定者。

以上任一例子,经由数据均可以得到一个可观测的样本比例。统计推断就是对相应的未知总体比例进行推论。这一章我们会介绍如何基于样本数据通过贝叶斯方法推断总体比例。因所举例子均是基于真实贝叶斯分析,从而与下一章参数讨论相比,本章内容更加具体详尽。

总体比例的贝叶斯统计分析已经超出了前一章所举的例子,其中对于未知总体比例我们只允许三个值,0.3、0.5和0.7。用 π 表示未知总体比例,只要真实存在且固定的 π 值未知,我们就可以将其视作一个变量。由于 π 可以取0到1间的任何数值,从而 π 是一个连续变量。用符号

π 表示比例并不常见，但用希腊字母表示总体参数是统计学的通则。

贝叶斯定理不论对连续变量还是离散变量都适用。与前例不同，这里我们用连续概率密度而不是将变量的概率分配给具体的变量值来表示连续变量的先验及后验分布。尽管相关数学运算可能比较复杂，但这里我们更偏重于应用而视一般理论为其次。因此，在本章中，从总体比例开始，我们就会给出具体范例和结果。

与之前离散例子相似，分析同样始于未知参数的先验分布。在指定了先验分布之后，收集数据，贝叶斯定理将作为一个桥梁将将指定参数的先验分布和所收集数据的信息连接起来以获得参数的后验分布。

一个例子

继续用之前关于宗教的例子，令 π 表示成年总体中天主教徒的比例。由于本身对 π 的认识非常有限无法给出相对精确的测量，但是可以肯定的是 $\pi<0.5$，且 π 应该位于 0.2 和 0.4 之间。基于这个模糊的认识，我画了一个密度图（见图 4.1）。

图中曲线和横轴之间的面积为 1，这意味着该图所展示的是一个概率密度分布，从而曲线以下的部分可解释为概率。可以发现，大多数的概率都是位于 0.1 和 0.5 之间，

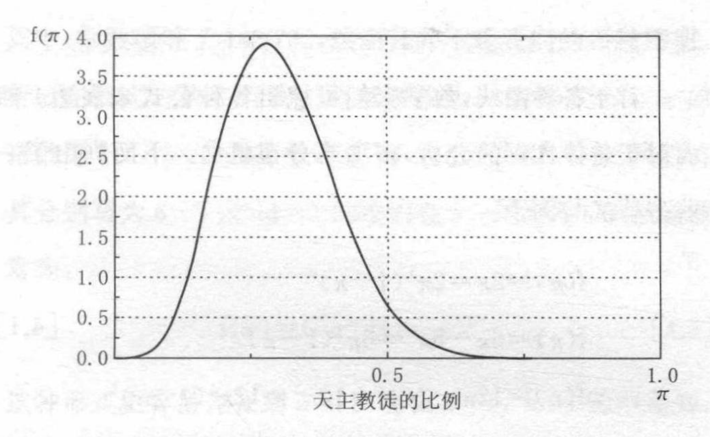


图 4.1 π 的先验分布,美国成年总人口中天主教徒的比例

其中大约 2/3 的概率位于 0.2 和 0.4 之间。因此,这张图与我们对 π 的初步认识非常吻合。

如前所述,贝叶斯分析并不要求所假定的未知参数先验分布必须是“正确的”。相反,其他人在做同样分析的时候可以假定一个非常不同的先验分布以表达参数的不确定性。此时,问题在于不同的先验分布是否会对最终结论有任何实质性的影响呢,我们将在第 6 章进行相应的讨论。

接下来,借用贝叶斯定理,我们将先验分布与数据信息合并。因 π 是一个连续变量,其可取 0 到 1 之间的任何值。对于前例,其背后的假设为:π 是离散变量,其仅包含了三个可能值,因此我们不能像之前处理单个数值一样处理 π 为连续的情况。一种处理连续变量的方法就是为图 4.1 中的曲线寻找一个合适的数学函数。该数学函数既包含曲线本身的信息又可用在处理连续变量的贝叶斯定

理中。

对于各种曲线，数学家们可想出各种公式来表达。然而对于总体比例的分析，多项式分布足矣。下面列出了一些多项式的例子：

$$
\begin{aligned}
f(\pi)&=2\pi=2\pi^1(1-\pi)^0\\
f(\pi)&=6\pi-6\pi^2=6\pi^1(1-\pi)^1\\
f(\pi)&=12\pi-24\pi^2+12\pi^3=12\pi^1(1-\pi)^2
\end{aligned}
\qquad [4.1]
$$

对每一个函数作图后发现，没有一个可以呈现图 4.1 中的曲线。如，第一个函数表示的是一条直线。

但是这三个函数属于同类，它们均属于多项式，且可以写成数字常量、π 的指数以及 $1-\pi$ 的指数三个部分的乘积。其中，常量是为了调整曲线与横轴之间的面积令其为 1.00，从而将一般函数转化为概率密度函数。而曲线的形状可以通过改变 π 与 $1-\pi$ 的指数来调整。当指数大于 0，曲线起始于原点，当 $\pi=1$ 时再次回到横轴，如图 4.1 所示。另外，指数越大，曲线的峰值越高。当 π 的指数大于 $1-\pi$ 的指数时，曲线峰值所对应的 π 值大于 0.5；当 $1-\pi$ 的指数大于 1，那么曲线峰值所对应 π 值小于 0.5。由图 4.1 中曲线形状可知：$1-\pi$ 的指数大于 π 的指数。

图 4.1 曲线的函数可以写为：

$$
\begin{aligned}
f(\pi)&=162\,792\pi^5(1-\pi)^{13}\\
&=C'\pi^{6-1}(1-\pi)^{14-1}
\end{aligned}
\qquad [4.2]
$$

其中，常数项等于 162 792，然而其并不是我们的兴趣所在，所以在公式的第二行我们将其替换为一般形式 C'。π 的指数等于 5，$1-\pi$ 的指数等于 13，在公式的第二行，我们将其分别写为 6－1 及 14－1 的原因在于，公式的一般形式通常为：

$$f(\pi)=C'\pi^{a-1}(1-\pi)^{b-1} \quad [4.3]$$

这种形式更容易先找到 a 和 b，那么 a－1，b－1 随即得知。最后一个函数为一个概率密度函数的一般形式，即贝塔分布(beta distribution)。该分布包含两个非负常数 a 和 b，以及一个取值在 0 到 1 之间的变量，当 a 和 b 都为整数时，等式 4.3 中常数 C'变为：

$$C'=\frac{(a+b-1)!}{(a-1)!(b-1)!} \quad [4.4]$$

值得注意的是，贝塔分布与二项分布非常不同。尽管两种分布的常量都是用阶乘表示，但是贝塔分布的 C'不能写成二项系数的形式。对于贝塔分布，分子中有 a＋b－1 个阶，而分母只有(a－1)＋(b－1)＝a＋b－2 个阶，而二项系数的分子和分母中所包含的阶数相同。另外，贝塔分布适用于取值为 0 到 1 间的连续变量分布，且其指数为常数，而在二项分布中，变量存在于指数中，且为离散型，取值范围为从 0 到 n。

有两种主要方法可以找到所需的贝塔先验分布，即，

找到可以确定分布的常量 a 和 b。第一种方法就是不断试错，即用电脑做出不同组合的 a、b 下的贝塔分布，选择一个最接近图 4.1 概率分布的 a、b 值作为先验知识。图 4.1 所表示的曲线为 a=6 和 b=14 的贝塔分布。由于与该 ab 组合临近取值下所对应的曲线不会与图 4.1 所示曲线有显著差异，因此也可以用其他曲线来表达有限的认知。

第二种方法是通过指定随机变量 π 的期望值和标准差来确定先验分布。π 的期望值为曲线分布的重心，这里认为分布在 $\pi=0.30$ 达到平衡点。由于大多数概率应位于 0.10 和 0.50 之间，因此认为该分布存在两个标准差的距离是合理的，即从均值 0.30 到 0.50，相差 0.20 的距离。若 0.20 表示两个标准差，那么一个标准差即 0.10。

贝塔分布的一个属性是通过常量 a 和 b 可以直接确定其均值和方差。同理，若知道了均值和方差，也很容易得到 a 和 b 的取值。这里令 μ 表示 π 的均值，σ 为 π 的标准差。当 π 以参数为 a、b 的贝塔分布时，均值、方差与 a、b 的关系为：

$$\mu=\frac{a}{a+b} \qquad \sigma^2=\frac{\mu(1-\mu)}{a+b+1} \qquad [4.5]$$

同样，我们也可以用均值和方差来表示 a、b：

$$a=\mu\left[\frac{\mu(1-\mu)}{\sigma^2}-1\right] \quad b=[1-\mu]\left[\frac{\mu(1-\mu)}{\sigma^2}-1\right] \qquad [4.6]$$

若已知均值和方差，通过等式 4.6 可找到确切的 a、b 的取值。

对于此例，我们已经确定 $\mu=0.30$，$\sigma=0.10$，从而 a、b 值也容易得出。且通过 a、b 值即可建立相应数学函数，图 4.1 中所示的曲线就可以画出了。

$$a=0.3\left[\frac{0.3(1-0.3)}{0.01}-1\right]=6$$

$$b=(1-0.3)\left[\frac{0.3(1-0.3)}{0.01}-1\right]=14$$

与此例一样，通过将均值和方差带入公式所得 a、b 并不一定总是整数。其实非整数取值也并非不可，只是需要将阶乘转化为伽马函数（gamma function）这个额外的步骤。大多数情况下，这些多余的努力并不一定会有显著的收益。从而，若计算出的 a、b 值非整，我们通常将其四舍五入。

至此，作为曲线和数学函数的先验分布就指定好了。现在我们需要的是通过其他信息来源即数据去了解 π。这里，所用的数据是一个包含 1 830 个人的随机样本，通过询问宗教信仰这样一个问题，共有 420 个样本报告了他们的天主教徒身份。

基于天主教徒在总体中的比例为 π，贝叶斯定理要求我们得到含有 420 个天主教徒和 1 410 个非天主教徒的概率。由于我们处理的是一个二分，概率为常量且观测间相

互独立的分布，因此可以通过二项分布公式得到概率。

$$\mathrm{Prob}(\mathrm{Data}\mid\pi)=\begin{bmatrix}1\,830\\420\end{bmatrix}\pi^{420}(1-\pi)^{1\,410} \qquad [4.7]$$

尽管参数 π 未知以及无法得到该数据的概率数值，但是所有关于 π 的信息均可通过等式 4.7 所计算得到。

根据贝叶斯定理，分子可以通过令等式 4.7 所得数据概率与等式 4.2 所得的先验概率相乘得到，

$$\begin{bmatrix}1\,830\\420\end{bmatrix}\pi^{420}(1-\pi)^{1\,410}\cdot 162\,792\pi^{5}(1-\pi)^{13}=C\pi^{425}(1-\pi)^{1\,423} \qquad [4.8]$$

其中常量 C 为二项系数与 162 792 的乘积。指数 425 和 1 423 分别通过 420 与 5，及 1 410 和 13 相加得到。

如上章讨论离散情况时所提及的，离散事件的分母可以通过将上式中三个乘积相加得到。而对于连续事件，分母是对 π 进行积分，其中积分区间为 0 到 1。然而，在两种情况下，分母均为一个常数，且其必须使得曲线下的总概率等于 1。

由于分母为一个常数，不论是何种取值，其均可与等式 4.8 中的常数合并生成一个新常数 C''。此时，π 的后验分布变为：

$$\begin{aligned}f(\pi\mid\mathrm{data})&=C''\pi^{425}(1-\pi)^{1\,423}\\&=C''\pi^{426-1}(1-\pi)^{1\,424-1}\end{aligned} \qquad [4.9]$$

该表达与等式4.2中π的先验分布的区别仅在于指数和常数的变化。因两者形式相同，所以π的后验分布也服从贝塔分布。通过等式4.3和等式4.4，我们知道，

$$C''=\frac{(426+1\,424-1)!}{(426-1)!\ (1\,424-1)!}=\frac{1\,849!}{425!\ 1\,423!}$$

所得C''非常大，除非我们想画出等式4.9中的曲线，否则对此无需过于关注。

由于π的后验分布服从贝塔分布，通过等式4.5，所得π的后验均值为：

$$\text{后验均值}=\frac{426}{426+1\,424}=\frac{426}{1\,850}=0.230$$

后验方差及标准差为：

$$\text{后验方差}=\frac{0.230(1-0.230)}{426+1\,424+1}=0.000\,095\,8$$

$$\text{后验标准差}=0.009\,8$$

图4.2展示了π的后验分布图。可以发现，大多数概率集中在0.21和0.25之间，从而，我们几乎可以确定π的取值也应落入此区间。更确切地说，即便曲线服从贝塔分布，因其足够对称，我们可将其近似为正态分布。由于π的标准差为0.009 8，那么，1.96个标准差等于0.019。用均值加上或者减去这个数值分别得到0.211和0.249两个值。因此，π位于0.211及0.249之间概率为0.95。该结果的正

式表达为：

$$\text{Prob}(0.211<\pi<0.249)=0.95$$

概率值 0.95 来自正态分布，其中当概率等于 0.95 时，标准正态分布 Z 值位于 −1.96 与 1.96 之间。

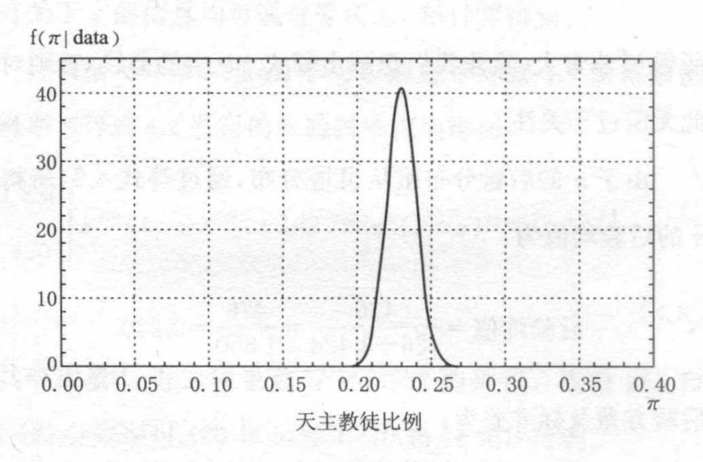


图 4.2 π 的后验分布

图 4.1 在一定程度上表达了最初对 π 先验分布的不确定性。通过样本数据，我们知道了更多关于 π 的信息，结合之前先验分布又进一步更新了对 π 的后验分布不确定性（如图 4.2）。通过两个分布的比较反映了 π 的不确定性的变化。之前我们认为 π 的取值大概在 0 到 0.60 之间，而现在我们有 95%的把握认为 π 的取值应该位于 0.21 和 0.25 之间。基于数据所带来的信息，将图 4.1 中相对分散的概率分布转化为图 4.2 中的尖峰分布。尽管曲线下的面积都为 1，但由于两分布纵轴的量度不同，很难直接进行比

较，只有统一了两张图的量度后才可比。图 4.3 展示了统一量度后 π 的先验和后验分布。

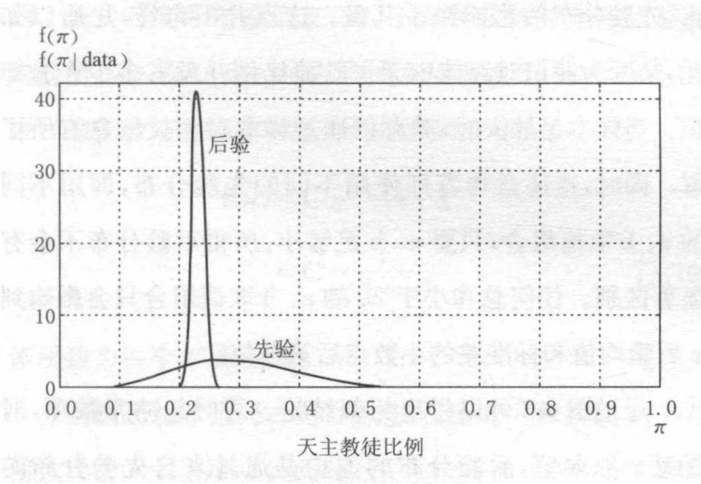


图 4.3　π 的先验和后验分布

通过 π 后验分布均值公式以及图 4.3 中的曲线，我们可以对 π 先验分布的影响有一定了解。以下为后验均值的计算：

$$\frac{426}{1\,850}=\frac{420+6}{1\,830+20}=\frac{\#\text{样本中的天主教徒}+\text{先验常数 a}}{\#\text{样本观测}+\text{先验常数}(a+b)}$$

通过样本所估计的 π 值等于样本比例 420/1 830＝0.230。如果我们将从贝叶斯分析中得到的 π 的后验均值作为贝叶斯分析中 π 的估计值，可以发现，相对于样本比例，分子增加了 6 而分母增加了 20。因此先验分布的影响可以看作，当样本量增加了 20，所增加的 20 个观测中有 6 个为天主教徒。先验知识可以用来推断 20 个观测样本，而

数据所提供的信息则可以用来推断 1 850 个观测样本。

这意味着后验分布主要由数据中的信息决定，相对而言，先验分布的影响微乎其微。这点并不奇怪，之所以如此，是因为我们对总体中天主教徒比例并没有多少先验知识。当样本足够大时，就可以通过样本对相关信息有所了解。同时，这还意味着即便用不同的先验分布，即用不同的 a、b 取值组合，只要 a、b 足够小，所得后验分布不会有显著区别。任何总和小于 20 的 a、b 取值组合只会影响到 π 后验均值和标准差的小数点后第三位。

通过图 4.3 可以得出相似结论。暂时忽略常数项，对任意 π 的取值，后验分布的值均是通过来自先验分布的 $\pi^5(1-\pi)^{13}$ 和来自数据的 $\pi^{420}(1-\pi)^{1410}$ 的乘积所决定的。当 π 小于 0.20 或者大于 0.26，后验分布及数据函数 $\pi^{420}(1-\pi)^{1410}$ 都几乎为 0。因此不论先验分布的 π 取何值对所得后验分布基本没有影响，原因在于先验分布所乘的数几乎为 0。从而当 π 的取值在 0.20 与 0.26 之间时，先验分布基本为常数，而后验分布差不多主要由该区间的数据函数决定。有关该问题在第 5 章会进一步讨论(稳健估计)。

无先验信息

问题在于先验分布可以被完全忽略吗？假设我们对总体中天主教徒比例完全没有概念，那么用 a＝6、b＝14

的先验分布与通过一个样本量为20，其中含有6个天主教徒和14个非天主教徒的样本作为我们的先验信息是一样的。若我们没有先验信息，即，不存在任何样本可以表达先验信息。那么，我们可以尝试常规假设用 $a=0$ 和 $b=0$ 来作为先验信息。此时所得到的先验分布为：

$$f(\pi)=C'\pi^{0-1}(1-\pi)^{0-1}$$
$$=C'\frac{1}{\pi(1-\pi)} \qquad [4.10]$$

该函数为一个U形曲线。

当缺乏总体比例的先验信息时，还有一种常用a、b取值选择为 $a=1$、$b=1$。此时，$\pi^{a-1}(1-\pi)^{b-1}$ 变为 $\pi^{1-1}(1-\pi)^{1-1}=1$，分布曲线形状为一条水平线。这就是广为所知的长方形分布或均匀分布（uniform distribution），即，π 的所有0到1间取值的概率都相同。严格地讲，一个连续变量取特定值的概率为0，该情况只适用于给定区间内 π 的取值概率。从均匀分布告诉我们，不论区间位置如何，只要其位于0到1之间，所得 π 的每一个取值概率都相同。

有关可能的完全忽略问题会在第6章信息先验与非信息先验一节中进一步讨论。

小样本

当样本量很小时，先验分布就会变得更加重要。因为

若小样本包含的信息非常有限,这时数据对于后验分布的重要性就会大大下降。因此,当数据有限的时候,尽可能得到更多的先验信息是很重要的。

研究数据和先验信息的影响的一种方法是看 π 的后验标准差的分母。因为若 a、b 值及数据取值范围较广,分子的值几乎恒定,除非我们研究的是当 π 值接近 0 或 1 的情况。而分母等于根号下 a+b 加上样本量再加 1。其中,先验信息的贡献为 a+b,样本的贡献为样本量。因此,与其说是一个样本量多大的问题,不如说是样本量 n 相对于 a+b 的大小问题。如上例,当 a+b 相对于样本量的比值较小时,a+b 的影响较小,其数值并不重要;而当 a+b 与样本量的比值接近 1 时,先验分布和数据的影响几乎同样重要。这个问题我们会在第 6 章先验分布的影响一节进一步讨论。

与经典方法的比较

经典统计学家可以通过参数估计,如点估计或者置信区间(confidence interval)告诉我们 π 有多大。对于我们的例子,点估计 p 等于样本中天主教徒的个数除以样本量,从而 $p=420/1\,830=0.229\,5$。一个被广为运用的贝叶斯点估计即后验均值。精确到小数点后第四位,该值等于 0.230 3。从数值上看,两者几乎一样。

π 的经典 95%置信区间可由二项分布的正态近似得到，并通过计算 $p \pm 1.96\sqrt{p(1-p)/n}$ 得到。将数值代入后，所得区间为 0.210 2 至 0.248 8。当精确到小数点第四位后，相应的 95%贝叶斯概率区间为 0.211 1 至 0.249 5。在实际研究中，我们很少做到如此精确，而这里我们这样做的原因在于，只有精确到小数点后第四位时，不同方法下的区间区别才得以显现。可以看出，贝叶斯区间相对于置信区间稍微靠右，而且较短。两种区别主要在于贝叶斯分析中先验分布的影响。

在一般情况下，因为先验分布基于有限信息，所以经典置信区间与贝叶斯概率区间在数值上非常接近，然而两者在解释上却存在着概念上的差异。对于置信区间，我们说如果有更多的样本并计算出每个样本的置信区间后会发现，在所得的所有区间中，有 95%包含未知 π，而剩余的 5%不包含 π。但是对于某个区间，如 0.21 到 0.25，我们并不知道其包含 π 与否。理想的情况是该区间属于一个更大的区间集，其中每个区间都包含 π。贝叶斯分析却不同，它允许我们通过先验分布表达最初对 π 的不确定性。在已知数据之后，尽管对 π 仍不确定，但是不确定性会随着对数据的了解逐渐降低，基于数据的新信息与先验分布所得的后验分布就可以对该不确定性进行更新。我们知道，参数的不确定性可以通过概率表达，而贝叶斯概率区间就是表达该不确定性的一种方法。对于上例，如用贝叶斯方

法解释就有，π 位于 0.21 与 0.25 之间的概率为 0.95。然而许多置信区间的使用者常常用贝叶斯推断解释一个置信区间，对于经典统计推断，这种解释是不可能的。

形式理论(formal theory)

作为总结，当指定未知总体参数的先验分布服从参数为 a 和 b 的贝塔分布时，我会介绍贝叶斯比例分析的形式理论。其中，先验分布的表达式为：

$$f(\pi)=\frac{(a+b-1)!}{(a-1)!(b-1)!}\pi^{a-1}(1-\pi)^{b-1} \quad [4.11]$$

先验均值 μ' 和方差 σ'^2 可以通过下式得到：

$$\mu'=\frac{a}{a+b} \quad \sigma'^2=\frac{\mu'(1-\mu')}{a+b+1} \quad [4.12]$$

同样，如果我们可以指定先验均值 μ' 和标准差 σ'，那么参数 a 和 b 即可根据下式计算出：

$$a=\mu'\left[\frac{\mu'(1-\mu')}{\sigma'^2}-1\right]$$
$$b=(1-\mu')\left[\frac{\mu'(1-\mu')}{\sigma'^2}-1\right] \quad [4.13]$$

假设数据满足二项分布，从而对于真实 π，数据的概率等于：

$$f(x|\pi)=\frac{n!}{x!(n-x)!}\pi^{x}(1-\pi)^{n-x} \qquad [4.14]$$

其中，n 为样本量，x 为我们所研究的具有某种或某些特征的观测值，而 $n-x$ 为其余的观测值。

通过贝叶斯定理可以将先验分布和数据概率结合起来，从而得到后验分布：

$$f(\pi|x)=\frac{(n+a+b-1)!}{(x+a-1)!(n-x+b-1)!}\pi^{x+a-1}(1-\pi)^{n-x+b-1} \qquad [4.15]$$

所得概率服从参数为 $x+a$ 和 $n-x+b$ 的贝塔分布。其后验均值 μ'' 及方差 σ''^2 分别等于：

$$\mu''=\frac{x+a}{n+a+b} \quad \sigma''^2=\frac{\mu''(1-\mu'')}{n+a+b+1} \qquad [4.16]$$

当指数非常大时，贝塔分布可以近似为正态分布，π 的贝叶斯概率区间可以通过 $\mu\pm z\sigma$ 得到，其中 z 为一个适当的标准正态分布变量值。如图 4.1 中的曲线，假设 $\mu'=0.30$，$\sigma'=0.10$，且 $z=2.00$，那么贝叶斯概率区间则为 0.10 到 0.50。当 $\pi=0.5$ 时，函数 $\pi^5(1-\pi)^{13}$ 等于 0.000 003 8；而当 $\pi=0.1$ 时，相应函数值为 0.000 002 5，这隐含了曲线并非完全对称，但是正态近似仍旧可行。

有关贝叶斯概率区间的精确计算可以通过贝塔分布和二项分布的相关性得出，如：

$$\text{Prob}(\pi < \pi_0) = \sum_{x=a}^{a+b-1} \begin{bmatrix} a+b-1 \\ x \end{bmatrix} \pi_0^x (1-\pi_0)^{a+b-1-x} \qquad [4.17]$$

例如，

$$\text{Prob}(\pi < 0.1) = \sum_{x=6}^{19} \begin{bmatrix} 19 \\ X \end{bmatrix} 0.1^x (1-0.1)^{19-x} = 0.009$$

$$\text{Prob}(\pi < 0.5) = \sum_{x=6}^{19} \begin{bmatrix} 19 \\ X \end{bmatrix} 0.5^x (1-0.5)^{19-x} = 0.968$$

从而，$\text{Prob}(0.1<\pi<0.5)=0.968-0.009=0.959$。对于相同区间，用正态近似所得概率为0.954，两者非常接近。

作比例分析时，贝塔分布并不是唯一可用的先验分布函数。尽管通过指定不同的a、b值，贝塔分布确实可以产生大量不同的曲线，但是仍可想象其他函数来指定 π 的先验分布。一种方法是把0到1的取值范围分成许多个区间，并通过先验分布寻找每个区间内 π 存在的概率。比如，将各区间的中点作为 π 的一个取值，此时，我们可将 π 视为离散型变量，如第3章所提到的第一个例子（即，假设 π 仅包含3个值），运用贝叶斯定理求得后验分布。

有关比例分析的更深入讨论可参见博克斯和蒂奥（Box & Tiao，1973）及施密特（Schmitt，1969）。

第5章

其他参数的贝叶斯方法

这章主要介绍有关均值、相关性、回归系数、列联表、均值差异、方差比率以及方差分析的贝叶斯方法。每一节都基于一个简单案例，并对如何获得参数的先验分布以及当数据满足某些假设时如何给出参数的后验分布进行讨论。

第1节 | 均值

坎贝尔和康弗斯(Campbell & Converse, 1980)在关于社会指标的项目中,对美国人进行抽样并让抽中的人对自己的生活标准进行评估,量度为0—100。这些数据可以用来研究总体中该变量的均值μ。

用贝叶斯方法对总体均值进行分析首先要通过先验分布表达对均值真值的不确定性。对于给定均值μ,如果我们知道来自某总体的数据的理论分布,那么就可以找到对应的数据概率。贝叶斯定理合并了先验分布的信息以及数据的信息,使得μ的后验分布容易算出。后验分布包含所有关于μ可用信息,其还可以用来寻找未知μ的点估计及区间估计。

平坦先验分布

数学理论已经解决了有关μ的两个不同先验分布,即,均匀分布和正态分布,我们先讨论均匀分布的例子。

均匀分布，又名长方形分布。顾名思义，根据该分布画出的图形状酷似长方形。该例的图形为横轴跨度为0—100，高度为1/100的长方形，这样的取值使得曲线下方的面积等于1。均匀分布令落入任何长度区间，范围在0—100之间的μ的概率都相同。从而，我们常说对于均匀分布，μ的任何取值的可能性都相同。

此例中使用μ的均匀先验分布表明我们对人们如何评价自己的生活标准知之甚少。我们将得到平均而言人们从非常不满意(0)到非常满意(0)之间的概率，而且均值可能落入这两个极值之间的任何位置。

已知数据中，生活标准这个变量包含$n=3611$个观测值。尽管数据直方图有些许偏斜，然而其与正态分布并无显著差别。观测样本均值$\bar{y}$等于76.91，样本标准差s等于18.46。

μ的均匀先验分布取值范围为负无穷到正无穷。根据贝叶斯定理，我们结合σ先验与服从正态分布的数据可以得到μ的后验分布。该后验分布服从自由度为$n-1$的t分布，其中μ的均值等于$\bar{y}$，标准差等于$s/\sqrt{n}$。基于该结果，即使μ取值范围仅为0—100，仍可得：

$$t=\frac{\mu-\bar{y}}{s/\sqrt{n}}=\frac{\mu-76.91}{18.46/\sqrt{3611}}=\frac{\mu-76.91}{0.31} \quad [5.1]$$

未知参数μ的均值，即，样本均值，等于76.91，μ的标准差

为 $18.46/\sqrt{3611}=0.31$。由于标准差如此小,因此无论 μ 的取值为负无穷到正无穷或是 0 到 100,其实都没有区别。例如,μ 的后验分布服从自由度为 3610 的 t 分布,拥有这般高的自由度,从而我们可用正态分布来代替 t 分布。

通过贝叶斯分析得出结论:总体中,美国人自我报告的生活标准的均值接近 76.91。更确切地说,美国总体均值有 0.95 的概率位于 $76.91-1.96\cdot 0.31=76.31$ 和 $76.91+1.96\cdot 0.31=77.51$ 之间。其他概率区间可以通过用正态分布的其他百分位数替换 $z=1.96$ 获得。

值得注意的是,该分析需要基于数据变量标准差的其他假设。因为这里我们假设变量服从正态分布,所以数据分布特征需要两个参数来标识,一个参数为均值 μ,对于 μ,其先验分布是基于均匀分布假设的;另一个参数为标准差 σ,对于 σ,我们还需知道它的先验分布以作进一步分析。

所得 μ 的后验分布服从 t 分布,其基于假设:对数标准差的先验分布服从均匀分布,且 μ 与 σ 是相互独立的。在大多数情况下,这是个非常合理的假设。更多有关该点的讨论请参考杰弗里斯(Jefferys, 1961)和施密特(Schmitt, 1969)。

最后,如果已知数据变量的标准差 σ 的数值,则无需将 σ 视为随机变量,也就无需假设 σ 的先验分布。这种情况下,由于均值 μ 先验分布服从均匀分布,后验分布则服从正态分布而非 t 分布,其后验均值和标准差如前所述。

正态先验与已知 σ

并不是所有人都把 μ 的先验分布用均匀分布来表示。如果有人有更多关于生活标准感知的信息,则可以将这些信息包含在先验分布中。令先验分布和数据均服从正态分布,且标准差已知,那么所得总体均值的后验分布则也服从正态分布。从而如果可行的话,当数据服从正态分布时,我们就可以试图用正态分布来表达先验知识。

如果一般来讲,人们对生活标准的感知大约为 50,即,两个极值的中心,那么总体均值可能大于 50 或者小于 50。更确切地说,假设 μ 的先验分布可以由一个均值为 50,标准差为 15 的正态分布表示,那么两个标准差就等于 30,从而我们几乎可以确定总体均值就位于 20 和 80 之间。

该分析需要我们知道总体的方差。然而,当样本量足够大时,就可以用样本方差 s^2 来表示总体方差 σ^2。

若 μ 和数据均服从正态分布,μ 的后验分布也服从正态分布。后验分布的均值则等于先验均值 μ' 与样本均值 $\bar{y}$ 的加权均值,其中权重为先验方差 σ'^2 和样本方差 σ^2/n 的倒数和。用数学符号表示为:

$$\text{后验均值} = \mu'' = \frac{\frac{1}{\sigma'^2}\mu' + \frac{1}{\sigma^2/n}y}{\frac{1}{\sigma'^2} + \frac{1}{\sigma^2/n}} \qquad [5.2]$$

将数值代入，我们有：

$$\mu''=\frac{\frac{1}{15^2}\cdot 50+\frac{1}{18.46^2/3611}\cdot 76.31}{\frac{1}{15^2}+\frac{1}{18.46^2/3611}}$$

$$=\frac{0.0044\cdot 50+10.5965\cdot 76.31}{10.6009}$$

$$=76.30$$

用方差倒数作为权重时要注意，当方差大时所得权重非常小，而当方差小时，所得权重很大，原因在于小方差暗示了均值变异性小，那么我们对所估计均值的取值确定性更强；相反，大方差表示，我们对所估计均值的取值不确定性更大。这种情况下，后验均值几乎与样本均值相等。因此，几乎所有的信息都来源于数据而不是先验分布。从而，样本均值的权重就要比先验均值的权重更大。

可以发现，该例一个重要的特征是样本量非常大。如前所提及，先验分布仅影响到后验均值小数点后第二位，即便数据方差 σ^2 的影响也几乎可以忽略。对于 $\sigma=18.46$，即，观测样本的标准差为 18.46，相应后验均值等于 76.30。换用不同的 σ，所得结果几乎相同。如，对于 $\sigma=10$，后验均值为 76.31，而对于 $\sigma=26$，所得后验均值为 76.29。因此，有关数据方差已知的假设并不重要，尤其当样本量非常大时，后验均值几乎不随数据分布的标准差的变化而变化。

μ 的后验方差可以通过结合 μ 的先验方差和样本均值

的方差得到，如下式：

$$\sigma''^2 = \frac{1}{\frac{1}{\sigma'^2} + \frac{1}{\sigma^2/n}} \qquad [5.3]$$

后验方差为方差倒数之和的倒数，其是两个方差的调和平均数(harmonic mean)的 1/2。代入数值后，我们有：

$$\sigma''^2 = \frac{1}{\frac{1}{15^2} + \frac{1}{18.46^2/3611}} = 0.094\,3$$

因此，μ 的后验标准差等于 0.307 1。通过等式 5.1，我们知道样本均值的标准差为 0.307 2，而 μ 的先验标准差为 15。如果不存在先验信息，那么 μ 的后验标准差则与样本均值的标准差相等。0.307 2 与 0.307 1 的区别仅在于先验信息的差异。由于从 0.307 2 到 0.307 1 变化很小，这从一定程度上反映了我们对总体均值的先验知识非常少。

有关总体均值的贝叶斯分析的更多讨论请参考，如菲利普斯(Phillips, 1974)和施密特(Schmitt, 1969)等书籍。

经典分析

这里，我们将样本均值视作随机变量，且服从 t 分布。经典统计分析将已知样本均值 76.91 作为未知总体均值 μ 的点估计，样本均值的标准误等于 $s/\sqrt{n} = 0.31$，因此总体

均值的 95%置信区间位于 76.91－1.96・0.31＝76.31 与 76.91＋1.96・0.31＝77.51 之间。基于双尾检验，我们则无法拒绝零假设，即，μ 等于 76.31 与 77.51 间的任何值。

一般来说，均值的置信区间与由均匀先验分布得到的贝叶斯概率区间在数值上相等。由于先验分布基于的信息更多，从而贝叶斯概率区间相较置信区间更小些。

即使区间在数值上相同，两者的解释在概念上截然不同。贝叶斯区间是以概率量度不确定性，它通过传达 μ 位于 76.31 与 77.51 之间的概率为 0.95 的信息来描述我们对总体参数的不确定性。而经典置信区间是基于概率的一个长期的相对频率。因此，它表明来自许多不同样本的置信区间中有 95%的区间包含总体均值，而剩下的 5%不包括总体均值。然而，我们并不知道 76.31—77.51 的特定区间是否包含总体均值。

有大量的证据表明很多研究者会私自把置信区间当做贝叶斯概率区间来解释。他们的理由是，由于来自大多数样本的区间确实包含未知总体均值，某一个特定区间包含真实参数的几率非常大。因此，他们将置信区间这个长期相对频率的概念替换为个人对不确定性的量度，即，以贝叶斯的概率区间概念来解释经典置信区间。但是贝叶斯推断与经典推断在概念上是无法共同成立的，在分析中，我们只能选择其一，而不能随便在两者间游移。在一定程度上，贝叶斯统计的吸引力就显而易见。

第2节 | 相关性

根据定义，相关性系数的数值位于-1.00与1.00之间。从而，相关系数的先验及后验分布落在区间外的概率应该为0。这种情况下，取值范围为负无穷到正无穷的正态分布或者t分布将无法成为相关系数先验及后验分布。再者，在数学上很难找到适用于贝叶斯统计和经典统计的精确的相关性系数分布形式，因此，我们常常用其转化形式或其近似形式进行分析。

令总体相关系数ρ的先验分布服从取值范围为-1.00到1.00的均匀分布，这意味着未知ρ可存在于该范围内的任何位置。假设数据服从二元正态分布(bivariate normal distribution)，n对观测值会有一个相关性系数r。若ρ先验服从均匀分布，通过贝叶斯定理，结合ρ的先验分布与正态分布数据所得后验分布就很难处理，从而我们用ρ的转换形式来进行分析。

这里，根据等式5.4，我们将ρ转化为新变量ζ：

$$\zeta=\frac{1}{2}\ln\frac{1+\rho}{1-\rho} \qquad [5.4]$$

其中，ln 表示自然对数。对于 ρ 的中间值，其与 ζ 没有区别。例如，若 ρ 位于 -0.50 与 0.50 之间，相应的 ζ 则位于 -0.55 与 0.55 之间。但是随着 ρ 越来越接近两个端点，将其 $+1$ 或者 -1 后，相应 ζ 的取值变为正无穷或负无穷。

将均匀先验分布与由二元正态分布生成的数据结合，所得 ζ 的后验分布变为均值为 μ''，方差为 σ''^2 的正态分布，其中，

$$\mu''=\frac{1}{2}\ln\frac{1+r}{1-r}-\frac{5r}{2(n-1)}\quad \sigma''^2=\frac{1}{n-1} \qquad [5.5]$$

根据上式，均值和方差均可以算出来，且能用来确定 ζ 的贝叶斯概率区间。这些区间能够由 ζ 转化为 ρ，因而也就得到了 ρ 的概率区间，如下例。

一个例子

在一个 $n=50$ 的随机样本中，我们得到两个变量关系的强度即相关系数为 $r=0.60$。如果未知总体中两个变量相关系数的先验分布服从 -1 到 $+1$ 的均匀分布，那么我们从后验分布中能获得 ρ 的什么结论呢？

根据以上均值和方差公式，我们知道被转化了的 ζ 的后验分布服从正态分布，其均值为：

$$均值=\frac{1}{2}\ln\frac{1+0.6}{1-0.6}-\frac{5\cdot 0.6}{2(50-1)}=0.693-0.031=0.662$$

以及

$$方差=\frac{1}{50-1}=0.020\,4$$

$$标准差=0.143$$

ζ 的95%贝叶斯概率区间为从 $0.662-1.96\cdot 0.143=0.382$ 到 $0.662+1.96\cdot 0.143=0.942$。若要对总体相关性系数的特征进行总结，我们还需将 ζ 变为 ρ。ζ 区间的端点需从 ζ 量度变为 ρ 量度，通过等式 5.4 就可以实现该步骤：

$$\rho=\frac{e^{2\zeta}-1}{e^{2\zeta}+1} \qquad [5.6]$$

从数值上来看，当 $\zeta=0.382$ 时，我们有：

$$\rho=\frac{e^{2\cdot 0.382}-1}{e^{2\cdot 0.382}+1}=\frac{2.147-1}{2.147+1}=0.36$$

类似地，当 $\zeta=0.942$ 时，我们有：

$$\rho=\frac{e^{2\cdot 0.942}-1}{e^{2\cdot 0.942}+1}=\frac{6.580-1}{6.580+1}=0.74$$

ρ 的点估计为后验分布的均值。ζ 的后验均值等于 0.622，将其转化为 ρ，其相应的值为 0.58。那么，ρ 的区间估计为 0.36 到 0.74，也就是说，ρ 落在该区间的概率为 0.95。

然而，需注意的是该区间非常大。有人可能认为在样本量为 50 的情况下，也许可以较为精确估计总体相关性系

数，然而我们可以确定的只是 ρ 位于 0.36 到 0.74 之间的某处。

有关贝叶斯分析中，均匀先验分布相关性系数的详细讨论，可参见博克斯和蒂奥（Box & Tiao, 1973）。

更多信息先验（informative priors）

我们可以知道的不仅限于总体相关系数的取值区间，即，相关系数位于－1 到＋1 的区间的某处，还可以得到更多有关总体的相关性信息。由于较早的研究可能已经对一些关于 ρ 的信息有所涉及，基于此，我们可以借以指定具有更多信息的先验分布。通过将其转化为正态分布，借用均值分析的一些结果可以得到 ρ 的后验分布。尤其在小样本的情况下，先验信息尤为重要。

继续之前的例子，假设先验知识可以让我们几乎确信 ρ 的取值在 0.20 和 0.80 之间。通过等式 5.4，将 ρ 的两个端点值转化为相应的 ζ，我们有：

$$\zeta_L = \frac{1}{2}\ln\frac{1+0.20}{1-0.20} = 0.203$$

$$\zeta_U = \frac{1}{2}\ln\frac{1+0.80}{1-0.80} = 1.099$$

其中，L 和 U 变为表示下限和上限。

由于 ζ 是一个正态分布变量，令 ζ_L 和 ζ_U 为以均值为

中心,对称的两个尾端的值,这样取值就可几乎涵盖两端点间所有的概率。从而,ζ 的先验均值就位于以两值为端点的区间中点,它们之间约有 4 个标准差。此时,先验分布的均值等于:

$$\mu' = \frac{\zeta_U + \zeta_L}{2} = \frac{1.099 + 0.203}{2} = 0.651 \qquad [5.7]$$

先验标准差等于:

$$\sigma' = \frac{\zeta_U - \zeta_L}{2} = \frac{1.099 - 0.203}{4} = 0.224 \qquad [5.8]$$

若数据为包含 50 个观测的随机样本,其中样本相关系数 $r=0.60$。将其转化为 ζ 量度,并用 z 表示所观测的样本值,我们有:

$$z = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\frac{1+0.60}{1-0.60} = 0.693$$

将观测到的相关系数转化为 z,随机变量 z 则服从正态分布,其方差为,

$$\text{var}(Z) \approx \frac{1}{n-3} = \frac{1}{50-3} = \frac{1}{47} = 0.021\,3$$

对于 ζ 量度,我们现有一个均值为 $\mu'=0.651$,方差为 $\sigma'^2=0.224^2=0.050$ 的正态先验分布。数据包含一个观测 $z=0.693$,且该观测来自方差为 $\sigma^2=0.021\,3$ 的正态分布。从而,ζ 的后验分布也是正态分布,其均值 μ'' 可通过等式

5.2 计算得出，

$$\mu''=\frac{\frac{1}{0.050}0.651+\frac{1}{0.021\,3}0.693}{\frac{1}{0.050}+\frac{1}{0.021\,3}}$$

$$=\frac{20\cdot 0.651+47\cdot 0.693}{20+47}$$

$$=\frac{45.591}{67}$$

$$=0.682$$

方差 σ''^2 可通过等式 5.3 得到，

$$\sigma''^2=\frac{1}{\frac{1}{0.050}+\frac{1}{0.021\,3}}=\frac{1}{20+47}=0.014\,9$$

从而，后验标准差 $\sigma''=0.122$。

95%的 ζ 贝叶斯概率区间范围为从 $\zeta_L=\mu''-1.96\sigma''$ 到 $\zeta_U=\mu''+1.96\sigma''$。将数值代入，得到 $\zeta_L=0.44$，$\zeta_U=0.92$。

最后，根据等式 5.6 将区间转化为 ρ 量度，得到：

$$\rho_L=\frac{e^{2\cdot 0.44}-1}{e^{2\cdot 0.44}+1}=\frac{1.41}{3.41}=0.41$$

$$\rho_U=\frac{e^{2\cdot 0.92}-1}{e^{2\cdot 0.92}+1}=\frac{5.30}{7.30}=0.73$$

因此，ρ 取值在 0.41—0.73 之间的概率为 0.95。

结合均匀先验分布，所得 95%后验概率区间为 0.36—0.74，若得到更多的先验信息，该区间变为 0.41—0.73。尽

管变化不是特别大，但是它足以表现出得到更多信息可以对先验分布产生影响。这里，ρ 先验分布没有产生很大影响的原因在于这里仅仅指定了 ρ 最可能位于 0.20—0.80 之间，而该区间仍旧非常大，因此并不能表示很多先验知识。但是可以发现的是，尽管先验信息有限，95% 概率区间范围仍旧减少了 1/10。

第3节 | 回归

简单回归

当两变量之间的关系为线性时，回归直线的斜率可以告诉回归直线向上斜还是向下斜，有多陡。通过样本数据，可以计算出样本的斜率 b。且通过样本斜率 b 可进一步推断总体斜率 β 以更好地理解两变量间的相关性。

在贝叶斯推断中，我们根据贝叶斯定理，通过结合先验信息与数据中的信息计算 β 后验分布来推断未知 β。

但是 β 并不是线性回归模型中的唯一参数，我们还需考虑其他两个参数。一个是截距项 α，还有一个是残差项的标准差 σ。尽管我们可能只对其中一个参数感兴趣，但这三个参数必须同时包含在分析当中。

令 β 的先验分布在可能 β 取值范围内服从均匀分布。类似地，令 α 以及 σ 对数的先验分布在其可能取值范围内也服从均匀分布。因三个参数的真实值未知，从而将其视

为随机变量，并假设 α、β 和 σ 的对数彼此间相互独立。在收集数据之前若缺乏有关参数的信息时，这几个常用的先验分布即可派上用场。通过结合先验分布与服从正态分布的已知数据，很容易就可以得到对应的后验分布。

通常，我们假设所收集的数据服从正态分布。且对于自变量 X 的一个固定值 x，我们假设因变量服从均值为 $\alpha+\beta x$，方差为 σ^2 的正态分布。因此，对于不同的 X 值，因变量的均值落在一条直线上，且不论 X 取何值，围绕该均值线的方差为 σ^2。

通过数据我们可以计算观测到的样本回归线，其中该回归线的截距为 a，斜率为 b，残差方差为 s^2。

结合取值范围为负无穷到正无穷的均匀先验分布与服从正态分布的数据，我们可以通过贝叶斯定理找出 3 个参数的后验分布。尤其当未知 β 的后验分布服从自由度为 $n-2$，均值为 b 且方差为 $s^2/\sum(x-\bar{x})^2$ 的 t 分布时。因此，样本斜率 b 就等于概率分布的均值，其表达了在进一步分析数据之后参数 β 的不确定性。

常常，我们把 b 作为 β 的一个点估计，从而贝叶斯概率估计会落入以下区间内：

$$b-t\cdot s/\sqrt{\sum(x-\bar{x})^2}$$

$$b+t\cdot s/\sqrt{\sum(x-\bar{x})^2}$$

其中，百分位数值 t 是通过自由度为 $n-2$ 的 t 分布得来。

继续沿用之前相关性部分的例子，基于样本量为 50 的样本数据所估计得到的回归线可用下式表达，

$$y=10.22+1.34x$$

其中回归线的残差平方和为 3 167.43，残差标准差为 $s=\sqrt{3\,167.43/48}=8.12$。因 $\sqrt{\sum(x-\bar{x})^2}=31.50$，从而 β 的后验标准差等于 8.12/31.50=0.258。含有 48 个自由度的 $\pm 2.01t$ 值可以给出 95% 概率区间，即，β 位于 $1.34-2.01\cdot 0.258=0.82$ 与 $1.34+2.01\cdot 0.258=1.86$ 之间的概率为 0.95。若不取 0.95 作为概率值，就需代入其他的 t 值。

此时，相关参数的非信息先验分布问题又摆在我们面前，很巧的是，β 的贝叶斯后验概率区间与参数的经典置信区间在数值上相同。因此，即便我们用置信区间表示，根据贝叶斯规则，我们仍可以用概率形式解释区间。

截距 α 的后验分布也服从自由度为 $n-2$ 的 t 分布，其中，α 的均值等于已知截距 α，方差为 $s^2\sum x^2/n\sum(x-\bar{x})^2$。将数值代入可知，$\alpha$ 的均值等于 10.22，标准差为 1.46。因此，α 落入区间 $10.22-2.01\cdot 1.46=7.28$ 与 $10.22+2.01\cdot 1.46=13.16$ 的概率为 0.95。95%的经典置信区间在数值上与该贝叶斯概率区间相同。

更多有关用非信息化均匀先验分布在简单回归中进行贝叶斯统计推断的讨论，请见施密特(Schmitt, 1969)。

斜率的信息先验

当残差项在回归线附近以正态分布,且方差为 σ^2,那么样本斜率 b 服从方差为 $\mathrm{Var}(b)=\sigma^2/\sum(x-\bar{x})^2$ 的正态分布。因此,我们可以将回归分析结果看做是一个数据观测,所得斜率 b,方差为 $\mathrm{Var}(b)$ 的随机变量。

我们可以把总体斜率 β 的先验认识表达为一个均值为 μ'、方差为 σ'^2 的正态分布,从而,该斜率服从正态先验分布。数据中包含一个观测值(可观测到的 b),且该观测值由一个正态分布得来。若这样看待回归分析,就可以回到之前的均值分析并运用当时的分析结果。尤其当参数 β 的后验分布服从均值为:

$$\mu''=\frac{\frac{1}{\sigma'^2}\mu'+\frac{1}{\mathrm{Var}(b)}b}{\frac{1}{\sigma'^2}+\frac{1}{\mathrm{Var}(b)}} \qquad [5.9]$$

方差为:

$$\sigma''^2=\frac{1}{\frac{1}{\sigma'^2}+\frac{1}{\mathrm{Var}(b)}} \qquad [5.10]$$

的正态分布时。

这些结果可以通过等式 5.2 和等式 5.3 得到,且均是建立在数据方差,即,残差的方差为已知的假设条件上。

当该假设不被满足时，尽管不准确，我们仍旧可以利用由数据得来的方差估计来计算，至少可以为后验分布找些感觉。当样本量越大时，所估计的总体方差越准确。

继续之前的例子，这里令总体斜率的先验分布服从均值 $\mu'=2.00$，标准差为 $\sigma'=0.75$ 的正态分布，而非之前的均匀分布。这时，我们几乎可以确定 β 位于 0.5 与 3.5 之间。数据包含 $b=1.34$ 一个值，且所估计的标准差为 0.26。将数值代入后验均值和方差的方程中，这时得到的 β 后验均值为 1.41，后验标准差为 0.25。因为 β 的值有很大可能性位于 0.90 与 1.90 之间，与之前相比，β 的不确定性大大减小了。

若要理解已知方差为 b 的假设对该分析的敏感性，我们可以尝试采取不同的方差值。之前根据 b 的标准差为 0.26，已经得到了 β 的后验均值和标准差。如果取值为 0.30，那么相应的后验均值和标准差分别为 1.43 和 0.28。对于$b=0.22$，则相应的后验均值和标准差分别为 1.39 和 0.21。可以发现，后验均值无显著变化，标准差则不然。

多元回归

一个含有 k 个解释变量的多元线性回归模型的主要包含 $k+2$ 个参数，分别为一个截距，一个残差方差，以及 k 个变量的回归系数。理论上讲，贝叶斯分析需要所有 $k+2$

个参数的联合先验分布来表达参数间的从属关系。根据贝叶斯定理,这样一个多元先验分布与数据分布结合后产生参数的一个多元后验分布。为了得到某个参数的后验分布,所得联合多元后验分布需要通过数学变换以消除其他参数影响。

通过假设 $k+1$ 个系数及残差标准差的对数的先验分布服从均匀分布且相互独立,以及数据来自多元正态分布,所得的每个回归系数的后验分布遵循自由度为 $n-k$ 的 t 分布。其中,参数 β_i 后验分布的均值为观测样本系数 b_i,标准差即 b_i 的标准差。该标准差即通常回归程序包生成的结果,但是对于贝叶斯分析,我们将其作为 β_i 而非 b_i 的标准差。

对于多元回归的贝叶斯推断的细节以及更多的数学讨论可参见博克斯和蒂奥(Box & Tiao, 1973)。

第4节 | 列联表

对于列联表，我们很难借用贝叶斯推断方法来进行分析，其原因主要在于基于这些表所建的模型大多是非参数型的。经典分析方法主要通过计算卡方统计量，但是该方法是基于假设：样本数据来自变量之间是相互独立的总体中。因此，一个重要的实质性问题，即，变量间的关系强弱程度如何，而有关该问题的答案则需计算尽可能多的潜在相关性量度，如 phi 或者 lambda。

贝叶斯推断可以专注于以上任意一个相关性量度。结合量度的先验分布与数据的先验知识就可以生成参数的后验分布。该计划的主要困难在于缺乏合适的分布形式。例如，给定总体中的相关性为 lambda，很难判定在一个样本列联表中的数据概率是多少，而该数据概率却又是贝叶斯定理所需要的。

考虑到这些困难，我们主要讨论两种方法，为简化计算，进一步将范围限制到 2×2 表格中。第一个方法主要考虑了两个比例间的差别，而第二方法主要用以处理 phi 系

数的分布。

两比例间的差异

通常，我们可以将分析 2×2 列联表考虑为研究两比例间的差异。假设列联表的列表示性别(男、女)，行表示党别(民主党、共和党)，那么性别与政党间的关系可以通过研究民主党的比例在不同性别中是否有差异而知。

对于该问题的贝叶斯分析可以先从研究两个比例开始。假设第一组(男性)在真实总体中的比例等于 π_1，且 π_1 的先验分布服从参数为 a_1、b_1 的贝塔分布。另外，样本包含 n_1 个观测，其中有 x_1 个民主党员，n_1-x_1 个共和党员。对于另一组(女性)，相应的参数分别为 π_2、a_2、b_2、n_2 与 x_2。那么，π_1 和 π_2 的后验分布分别服从均值为 μ_1、μ_2，方差为 σ_1^2、σ_2^2 的贝塔分布，如等式 4.16 所示。

基于以上信息，可以知道 $\pi_1-\pi_2$ 的后验均值等于$\mu_1-\mu_2$，$\pi_1-\pi_2$ 的后验方差等于 $\sigma_1^2+\sigma_2^2$。由于每个样本的观测均不在少数，从而我们可以将 $\pi_1-\pi_2$ 的分布近似为一个正态分布，基于该近似，我们就可以构建 $\pi_1-\pi_2$ 的贝叶斯概率区间。

用具体数值举个例子，假设 π_1 和 π_2 的先验分布相同且对称，其中 $a_1=a_2=b_1=b_2=5$。样本数据中，包含 $n_1=50$ 个男性和 $n_2=50$ 个女性，其中男性中有 $x_1=40$ 个民主

党员，而女性中有 $x_2=20$。π_1 的后验分布服从贝塔分布，其均值等于：

$$\mu''_1=\frac{x_1+a_1}{n_1+a_1+b_1}=\frac{40+5}{50+5+5}=\frac{45}{60}=0.75$$

方差为：

$$\sigma''^2_1=\frac{\mu''_1(1-\mu''_1)}{n_1+a_1+b_1+1}=\frac{0.75\cdot 0.25}{50+5+5+1}=\frac{0.187\,5}{61}$$
$$=0.003\,1$$

同样，π_2 的后验分布也服从贝塔分布，其均值等于：

$$\mu''_2=\frac{x_2+a_2}{n_2+a_2+b_2}=\frac{20+5}{50+5+5}=\frac{25}{60}=0.42$$

方差为：

$$\sigma''^2_2=\frac{\mu''_2(1-\mu''_2)}{n_2+a_2+b_2+1}=\frac{0.416\cdot 0.583\,3}{50+5+5+1}=\frac{0.243\,1}{61}$$
$$=0.004\,0$$

通过均值和方差，我们知道 $\pi_1-\pi_2$ 的后验均值为 0.75－0.42＝0.33，后验方差为 0.003 1＋0.004 0＝0.007 1，那么相应的后验标准差即 0.084。因样本的样本量足够大，从而可以将 $\pi_1-\pi_2$ 近似看作服从正态分布，其中均值为 0.33，标准差为 0.084。且其 95％的概率区间为 0.33－1.96 • 0.084＝0.17 到 0.33＋1.96 • 0.084＝0.49。因此，男女间民主党派比例的真实差异有 0.95 的概率落于 0.17 与 0.49 之间。

Phi 的分布

当 2×2 列联表的期望频率足够大到可以计算经典拟合优度量度——卡方时，表中左上角单元格服从均值为 r_1n_1/n，方差为 $r_1r_2n_1n_2/n^3$ 的近似正态分布，其中 r_1 与 r_2 分别为行合计数，n_1 与 n_2 分别为列合计数，n 为观察值的总个数。因此对于表 5.1 中的例子，可观测的正态分布变量值等于 40，其方差等于 $\sigma^2=60\cdot40\cdot50\cdot50/100^3=6$。

表 5.1　性别和政党的假设数据

		性别		总计
		男	女	
政党	民主党	$x_1=40$	$x_2=20$	$r_1=60$
	共和党	10	30	$r_2=40$
	总计	$n_1=50$	$n_2=50$	$n=100$
		phi = 0.41		

因边缘固定，所以方差已知，从而数据中包含了一个已知方差的正态分布变量的一个值。该变量的均值服从均匀先验分布，其后验分布均值则与样本均值相同，且其方差等于 σ_2^2/n。由于这里只有一个观测，因此样本均值就是我们所观测到的值，方差等于 σ^2。

就这个例子而言，数量的后验分布为 x_1，它服从一个均值为 40，方差为 6 的正态分布。其 95％贝叶斯概率区间

为 $40-1.96\cdot\sqrt{6}=35.20$ 到 $40+1.96\cdot\sqrt{6}=44.80$。phi 的观测值为 0.41。若将 x_1 替换为区间最小值 35.20，且令边缘保持不变，所得 phi 值为 0.21；若将 x_1 替换为区间最大值 44.80，相应的 phi 值变为 0.60。因此，我们有 95%的把握认为，两个变量的相关系数 phi 值位于 0.21 到 0.60 之间。

最后，若我们有理由相信给定均值和方差的情况下，x_1 的先验分布服从正态分布，那么结合该先验信息与数据信息，所得后验分布也应为正态分布，其中均值等于加权后的先验均值与可观测的数据 x_1 总和，根据等式 5.2 和等式 5.3，其方差等于两者方差倒数和的倒数。

第5节 两个均值间的差异

很多情况下，我们要研究两个独立观测组之间的均值差异，例如，男性与女性在收入上是否有区别？新教徒与天主教徒在对待流产的态度上是否有差异？等等。这里，两个组别由一个两分名义变量定义。有关该名义变量与定距因变量间是否存在相关关系是我们想要解决的问题。

一种检验二分名义变量与定距因变量之间相关性的方法是，创建一个关于名义变量的虚拟变量，并用该因变量对虚拟变量做回归。由于虚拟变量的取值为0和1，因此回归线斜率表示两组别在因变量均值上的差异，那么我们就可以用以上回归分析中所讨论的方法来研究两个均值之间的关系。

以前研究两均值间差异最常见的方法是基于原理一进行方法开发。然而，所得方程与正态分布假设下的回归分析完全一致。在经典统计推断中，基于两个总体的方差来自什么类型的数据，我们需要区分三种情况，即：当两个方差已知的时候、当两个方差未知但是相等的时候，以及

当两个方差未知但是不等的时候。

方差已知

特殊情况下，当来自总体的两组观测样本的方差 σ_1^2 和 σ_2^2 已知时，我们就可以直接计算两总体均值的区别 $\delta=\mu_1-\mu_2$，那么所观测到的两样本均值的差异 d 服从正态分布，且其方差等于：

$$\mathrm{Var}(d)=\mathrm{Var}(\bar{y}_1-\bar{y}_2)=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2} \qquad [5.11]$$

其中 n_1 与 n_2 分别为两组观测的样本量。需提及的是，这种情况非常少见。

基于以上信息，之前提及的有关均值分析的各种结果就可以用了。若 δ 的先验分布服从均匀分布，所得后验分布则服从正态分布，由等式 5.11 可知，该分布的均值为 d，方差为 $\sigma_1^2/n_1+\sigma_2^2/n_2$。而若 δ 的先验分布服从均值为 μ' 方差为 σ'^2 的正态分布，那么 δ 的后验正态分布均值为 μ' 和 d 的加权平均，其中权重等于相应方差的倒数（见等式 5.2）。类似地，δ 的后验方差为方差倒数和的倒数（见等式 5.3）。

方差未知但相等

在埃尔德（Elder，1969）的有关奥克兰成长研究中提

到了这样一个问题，中产阶级和工人阶级的女孩在外表上有区别吗？数据样本中有35个女孩来自中产阶级家庭，43个来自工人阶级家庭，研究者对其身体素质进行了测量，结果发现，来自中产阶级的女孩身体素质均值为56.6，标准差为13.5，而来自工人阶级的女孩身体素质相应均值和标准差分别为48.6、14.2。

我们假设两总体的方差σ_1^2、σ_2^2相等，从而有$\sigma_1^2=\sigma_2^2=\sigma^2$，且$\mu_1$、$\mu_2$以及$\log\sigma$相互独立。令$\mu_1$、$\mu_2$和$\log\sigma$的先验分布服从均匀分布。结合正态分布数据，可以得到自由度为n_1+n_2-2，$\delta=\mu_1-\mu_2$的后验分布服从均值为$d=\bar{y}_1-\bar{y}_2$，方差为$s^2(1/n_1+1/n_2)$的t分布，其中s^2为将两样本合并后所得方差。

对于此例，将样本合并所得方差为：

$$s^2=\frac{(35-1)13.5^2+(43-1)14.2^2}{35+43-2}=\frac{14\,665.38}{76}$$

$$=192.965\,5$$

δ的后验方差从而等于192.965 5/(1/35+1/43)=10.000 9，这表明其后验标准差为3.16。δ的后验均值等于56.6−48.6=8.0，t分布包含76个自由度。因此，δ有0.95的概率落在8.0−1.99 · 3.16=1.71与8.0+1.99 · 3.16=14.29之间。另外，如果δ大于0，那么中产阶级女孩的总体均值大于工人阶级女孩。这表明，平均而言，来自中产阶级的女孩比来自工人阶级的身体素质更好。对于$\delta=0$

时，所对应的 t 值为：

$$t=\frac{0-0.80}{3.16}=-2.53$$

从而可知 t 大于 -2.53 的概率为 0.99。因此基于所收集的测量为准，在奥克兰，来自中产阶级的女孩的身体素质均值有 0.99 的概率比来自工人阶级的女孩的均值高。

贝叶斯概率区间为 1.71—14.29，其与 95%的置信区间在数值上相同。尽管两者在概念上有很大差别，但是由于两者数值上相等，且利用之前得到的均匀分布的贝叶斯概率区间和观测到的数据，这里我们可以将均值差异的经典置信区间用贝叶斯概率区间解释。

方差未知且不等

当两组观测的方差 σ_1^2、σ_2^2 不等时，t 分布就不再适用。

令 μ_1、μ_2、$\log\sigma_1$ 以及 $\log\sigma_2$ 与均匀先验分布相独立。第一个样本数据服从均值为 μ_1 方差为 σ_1^2 的正态分布，第二个样本数据服从均值为 μ_2 方差为 σ_2^2 的正态分布。分别看参数 μ_1 和 μ_2，通过之前有关单个均值的分析可知，它们的后验分布均服从 t 分布。

但是我们想知道 μ_1 和 μ_2 差异的后验分布，由于 μ_1 和

μ_2 的方差不同，从而两均值差异的后验分布并不服从 t 分布。我们可以用 Behrens-Fisher 分布来表示该后验分布，然后这个分布并不容易处理，因此，我们可以将 Behrens-Fisher 分布近似为 t 分布，值得注意的是，该近似伴随着$\mu_1-\mu_2$ 的后验方差以及自由度的修改。

将 Behrens-Fisher 分布近似为 t 分布需要找到两个数字：a 和 b。从而 $\mu_1-\mu_2$ 的后验分布就可以近似服从自由度为 b，均值为 $\bar{y}_1-\bar{y}_2$，方差为 $a^2(s_1^2/n_1+s_2^2/n_2)$ 的 t 分布。仍以奥克兰女孩身体素质为例，根据博克斯和蒂奥(Box & Tiao, 1973)，我们可以通过以下过程找到 a 和 b：

$$u=\frac{s_1^2/n_1}{s_1^2/n_1+s_2^2/n_2}=\frac{13.5^2/35}{13.5^2/35+14.2^2/43}=\frac{5.207}{9.896}=0.526$$

$$v=1-u=1-0.526=0.474$$

$$f_1=\frac{n_1-1}{n_1-3}u+\frac{n_2-1}{n_2-3}v=\frac{34}{32}0.526+\frac{42}{40}0.474=1.057$$

$$f_2=\frac{(n_1-1)^2}{(n_1-3)^2(n_1-5)}u^2+\frac{(n_2-1)^2}{(n_2-3)^2(n_2-5)}v^2$$

$$=\frac{34^2}{32^2\cdot 30}0.526^2+\frac{42^2}{40^2\cdot 38}0.474^2=0.017$$

$$b=4+\frac{f_1^2}{f_2}=4+\frac{1.057^2}{0.017}=71.6$$

$$a=\frac{b-2}{b}f_1=\frac{71.6-2}{71.6}1.057=1.027 \quad [5.12]$$

该计算过程明确列出了当两观测样本方差不同时，我

们仍需做的调整。代入数值后，$\mu_1-\mu_2$ 的后验分布可近似为一个自由度 $b\approx72$，均值为 $\bar{y}_1-\bar{y}_2=56.6-48.6=8.0$，方差为 $a^2(s_1^2/n_1+s_2^2/n_2)=1.027^2(13.5^2/35+14.2^2/43)=10.44$ 的 t 分布。所得结果与基于同方差假设的结果不同。但是由于两观测样本的方差 $s_1^2=13.5^2=182.25$ 与 $s_2^2=14.2^2=201.64$ 差别并不大，因此即便无显著区别也并不奇怪。

需注意的是，用经典统计我们则无法得到相似结果。

第6节｜两个方差的比率

有时，我们有许多理由来比较两个方差。例如，我们想知道，就某些变量而言，某组组内同质性强于另外一组。但是，更多的情况下，我们对两个均值间的差异更感兴趣，这时，就需要先比较方差来确定用哪种方法对均值进行比较。

我们的样本数据来自两个正态分布的总体，其均值分别为 μ_1 和 μ_2，方差分别为 σ_1^2 和 σ_2^2。令 μ_1、μ_2、$\log\sigma_1$ 以及 $\log\sigma_2$ 相互独立且其先验分布均服从均匀分布，根据贝叶斯定理，我们可以得到比率：

$$F=\frac{s_1^2/s_2^2}{\sigma_1^2/\sigma_2^2} \qquad [5.13]$$

该比率的后验分布为含有 n_1-1 和 n_2-1 自由度的 F 分布，其中 s_1^2 和 s_2^2 为两观测组的方差，n_1 和 n_2 为两观测组的样本量。

继续之前的例子，我们有：

$$F=\frac{13.5^2/14.2^2}{\sigma_1^2/\sigma_2^2}=\frac{0.90}{\sigma_1^2/\sigma_2^2}$$

方差比率的后验分布为含有 34 和 42 自由度的 F 分布。但是该结果并没有直接告诉我们有关 σ_1^2/σ_2^2 的后验分布，尽管该分布可能得到，有关这点请见博克斯和蒂奥（Box & Tiao，1973）。

另一个使用 F 分布得到 σ_1^2/σ_2^2 的贝叶斯概率区间方法是通过 F 分布表，我们可以找到两个值 F_L 和 F_U，使得 F 落在 F_L 和 F_U 之间的概率为 $1-\alpha$。换句话说，当概率等于 $1-\alpha$ 时，有：

$$F_L < F < F_U$$

其中，F_L 位于 F 分布的左尾，而 F_U 位于 F 分布的右尾。将 F 的公式代入，我们有：

$$F_L < \frac{s_1^2/s_2^2}{\sigma_1^2/\sigma_2^2} < F_U$$

解不等式，我们有：

$$\frac{s_1^2/s_2^2}{F_U} < \sigma_1^2/\sigma_2^2 < \frac{s_1^2/s_2^2}{F_L}$$

对于此例，

$$\frac{0.90}{F_U} < \sigma_1^2/\sigma_2^2 < \frac{0.90}{F_L}$$

我们所要做的就是找出 F_L 和 F_U。

如果我们想得到 95% 的概率区间，那么 $\alpha=0.05$。一种选择 F_L 和 F_U 的方法就是找出 F_U，使得 F 小于 F_U 时概率等于 0.975，然后找出 F_L，使得 F 小于 F_L 时概率等于

0.025。此时，F 位于 F_L 和 F_U 之间的概率等于 0.95。F_U（位于分布右尾的较大值）可以直接通过 97.5%F 表得出。对于自由度为 34 和 42 分布，我们得到 $F_U=1.89$，但是位于 F 分布左尾的值通常不会直接给出，因此还需要额外的步骤来找出 F_L。首先反转自由度，这时我们有 $n_2-1=42$ 有 $n_1-1=34$ 个自由度，然后在 97.5%表中查找相应的 F 值。F_L 则等于该值的倒数。对于自由度为 42 和 34 的分布，$F=1.94$，那么，相应的 $F_L=1/1.94=0.52$。

σ_1^2/σ_2^2 的 95%贝叶斯概率区间等于：

$$0.90/1.89 < \sigma_1^2/\sigma_2^2 < 0.52$$

$$0.48 < \sigma_1^2/\sigma_2^2 < 1.73$$

因此，我们对两个总体方差比率位于 0.48 和 1.73 之间有 95%的把握。

一个小问题

为得到 95%概率区间，在查找相应的 F_L 和 F_U 的时候，我们排除了 F 分布左尾的 2.5%和右尾的 2.5%。但是 F 分布本身是偏斜的，因为如此，从而区间 F_L 到 F_U 并不是我们期望的最短的 95%概率区间。比如说，如果我们排除了左尾的 2%和右尾的 3%，所得到的仍旧是 95%区间，但是该区间其实可以更小的。然后在大多数实践中，很多区间并无显著差别，从而我们无须在该细节上纠结过多。

第7节｜方差分析

“方差分析”包含了大量的统计模型用以分析一个或多个名义变量与一个定距变量间的关系。根据名义变量的数目，我们有单向，双向或多向分析。在经典统计分析中，我们还会区分随机效应模型和固定效应模型。贝叶斯统计中所有未知参数均被视为随机变量，即，假设随机效应模型和固定效应模型一样是不成立的。然而，通过对相关参数选择合适的先验分布，就有可能得出符合两类模型特征的方法对随机效应模型和固定效应模型加以区分。这里我们将分析限制到对多组均值比较进行单向分析，并从两均值的比较开始逐渐延伸。有关各种方差分析模型的讨论，请见博克斯和蒂奥(Box & Tiao, 1973)。

在艾弗森和诺尔波特(Iverson & Norpoth, 1976)有关方差分析的论文中，他们通过来自5个国家的主观政治能力测量构建了一个含有 $n=18$ 个观测的小假设样本。该五个国家的均值分别为4、6、2、7和5，可观测的 F 比率等于13.00，其中 F 分布自由度为4和13。基于经典统计

分析，结果发现1%显著性水平所对应的F值等于5.20。因为F的观测值大于5.20，从而总体的等均值假设被拒绝。

基于同样的数据，现在我们用贝叶斯方法来分析五个总体的均值是否相等。我们令五个总体均值和残差的对数标准差的先验分布服从独立均匀分布，并假设抽样数据来自未知均值但等方差的正态分布。根据贝叶斯定理，我们可以找到总体均值的联合后验分布。要研究五个总体均值是否相等，就需要通过后验分布构建一个联合贝叶斯概率集，使其包含五个总体均值的最可能值的集合，就如之前我们为单个参数构建贝叶斯概率区间一样。

因有关五个总体均值坐落在五个维度上，从而用联合贝叶斯概率集很难描述。但是对于两个总体均值μ_1和μ_2就有可能。图5.1中阴影部分为从μ_1和μ_2的联合后验分布获得的99%概率集。这意味着坐标(μ_1, μ_2)有0.01的概率会落于阴影部分以外。

在方差分析中，我们常常对总体均值是否相等感兴趣，即，$\mu_1=\mu_2$是否成立。该参数限制在μ_1、μ_2平面上定义了一个子集，且该子集为一条经过原点且斜率为1的直线，如图5.1中标注$\mu_1=\mu_2$的直线。任何经过该直线的点所对应的两个总体均值都相等。因图5.1中直线没有与概率集合相交，从而可知，两个总体均值不等的概率为0.99。

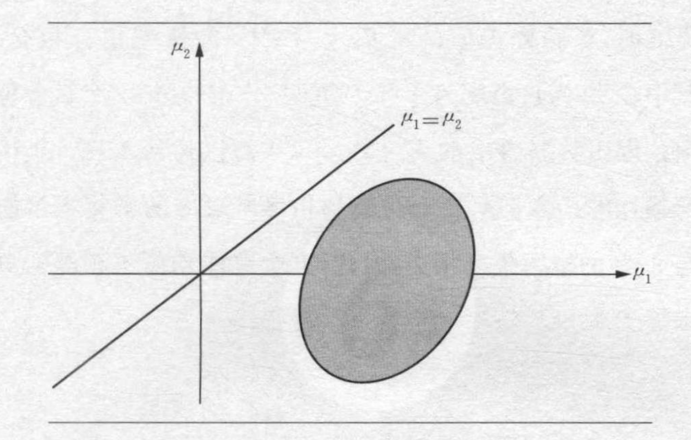


图 5.1　两个均值 μ_1 和 μ_2 的联合贝叶斯概率集(阴影部分)及其子集 $\mu_1=\mu_2$(直线)

回到之前五维情况，现在问题变为五个总体均值 μ_1、μ_2、μ_3、μ_4 和 μ_5 的联合后验贝叶斯概率集是否与所有均值都相等情况下的子集相交，即由 $\mu_1=\mu_2=\mu_3=\mu_4=\mu_5$ 所定义的子集。这个问题看似复杂，但是其完全可由观测到的 F 比率值所回答。

粗略得讲，只有在观测到的 F 值非常小时，等均值子集才会与可能参数值的贝叶斯概率集相交。同样，如果观测到的 F 很大，那么等均值子集则不会与概率集合相交。从而，一个小的 F 值才会令总体均值相等成为可能，而大的 F 值则暗示了所有均值都相等的情况概率很小。

更确切地说，当且仅当观测到的 F 值小于 F 分布的第 P 个百分位，等总体均值的子集才与 P%的概率集合相交。对于此例，令 P 等于 99，则有，当且仅当 F 的观测值小于

5.20 时，等总体均值的子集才与 99%的概率集合相交。其中 5.20 为自由度为 4 和 13 的 F 分布的第 99 个百分位值。因由数据得出的 $F=13.00$，从而该情形与图 5.1 中所描述的一样，即，五个均值都相等所对应的子集不可能与 99%的概率集合相交，因此，五个总体均值不相等的概率等于 0.99。

第6章

先验分布

在所有贝叶斯方法以及之前所举的例子中，我们都有用到先验分布这一概念。在本章，我们将会就该分布的各个方面进行更细致且系统的介绍。

本书第一次接触到先验分布是第3章讨论贝叶斯定理(用来区分三个不同的总体)的时候。很多时候确实存在不同的总体，究竟样本数据应该从哪个总体抽取则决定于概率机制，这就与掷骰子一样。范例中三个总体先验概率相等，均为1/3。在这种情况下，即使是古典统计学家，在数据已知的情况下，也会用贝叶斯定理来确定各个总体为数据来源的概率大小。但是当我们仅需处理一个总体且其参数未知时，只有贝叶斯推论会用到先验分布，经典统计推论却不存在类似概念。

在第一个例子中，先验分布包含一些信息，这些信息反映了一些对于总体的特定认知。特别是当概率不等且一个总体的概率比另一个高很多时，在收集数据之前，我们就会对总体如何生成数据有一个较好的了解。这与之

后一些例子不同，我们经常用均匀先验分布来反映我们有限的先验知识。

三个不同总体例子的第二个特征是先验概率容易得到。其先验概率是通过计算投掷一个公平的骰子可能产生的各种结果的概率得到的。在很多情况下，先验概率很难确定。例如，有时先验概率的确定是通过先验分布中参数的特定值得到的。例如，贝塔分布就是其中的一个，因对于该分布，我们必须指定参数 a 和 b 的值。

该例的第三个特征，即，先验概率的选择是非主观性的。当先验概率通过概率机制比如掷骰子确定，所有人都会得到同样的先验概率。由于先验概率不存在模糊性，与每个人有不同的先验概率分布相较，它们相对更"科学"。毕竟科学研究的结果不取决于研究是谁做的。然而对于不同人会有不同先验分布的情况，贝叶斯分析则并不将其视之为一个问题。

有关先验分布要讨论的最后一个问题，即，它对后验分布的影响有多大？在该例中，三个不同总体的先验概率相同，在一定程度上说明先验概率的几乎没有影响。原因在于它们在贝叶斯定理的分子分母中被消去了(见表 3.1)，从而后验概率与数据概率成正比。但是该例中的样本量非常小，其他不同的先验分布可能会得到非常不同的后验分布。

第1节 | 信息先验与非信息先验

本节会讨论两个相悖的论点。一个论点是我们研究问题是因为我们不知道问题的答案，通过研究我们可以探索新的未知领域。从而，研究的目的是为了创建一些未知参数的值，例如，回归直线的斜率可以帮助我们理解两变量间的关联性。此时，研究是用可能的参数值信息来填补之前未知参数值的空缺。

贝叶斯分析需要对未知参数假设一个先验分布。即使我们对这个参数并不了解，仍旧需要一个先验分布。这时就有两个问题，是不是存在“完全未知”，如果是，是否可以将其表达为一个先验分布。我个人认为在这个世界上，有很多东西对我来说都是完全未知的，有关这一点，我想这不仅仅是我一个人的感觉。

但是将完全未知表达为一个先验分布是另外一回事。一个人可以说，“我对这个回归系数一无所知。但是据我所知，系数的任何值都有可能，且不存在一个值比另一个更有可能”。言外之意，未知参数所有可能取值服从的是

均匀分布。但是即使是均匀分布,所引之言也包含了一些参数信息,即,每个取值表达参数信息的可能性都相同。因此,与完全无知相较,这也是一种进步。

均匀分布本身也隐含了多于完全无知的信息。一种证明方法就是反证法。假设一个取值为 0 到 5 的变量 X 服从均匀分布,那么 X 值小于 2.5 的概率等于 0.50。但如果我们对 X 完全无知,那么我们对 X^2 也完全无知。如果无知等价于均匀分布,那么 0 到 25 的范围内服从均匀分布就可以作为对 X^2 的无知。X^2 小于 12.5 的概率因此等于 0.50。若事实如此,那么 X 本身小于 $\sqrt{12.5}=3.54$ 的概率也应该等于 0.50。但是基于之前对 X 服从均匀分布的假设,X 小于 2.5 的时候概率才等于 0.50。因为概率等于 0.50 不可能在 X 小于 2.5 和 X 小于 3.54 时同时成立,基于 X 与 X^2 服从均匀分布的假设,这两个有关 X 的概率陈述相悖,所以不能用均匀分布来表示该完全无知的情况。

因完全无知本身并无意义,与其尝试任何有关完全无知的概念,不如试图区分信息先验与非信息先验。我们可以将非信息先验想象成所要研究的参数的最低先验认识。在大多数情况下,我们用均匀分布来表示这样一个先验分布。我们不试图表达完全的无知,这里所要表达的想法为,参数的所有值在相关参数的取值区间内的概率相同。更确切地说,一个固定长度的区间内的取值概率都相同,不管该区间位于参数取值范围的哪一段。

如我们在第 4 章、第 5 章所见的用贝叶斯方法分析的例子中，非信息先验所导致的结果是在数值上与通过经典统计分析所得结果相同。确切地说，贝叶斯概率区间与经典置信区间在数值上一致，然而在解释上，两种区间非常不同，但是两者在数值上的一致性表明了从经典统计推断到贝叶斯统计推断只需简单一步，即假设先验分布服从均匀分布。

一般说来，反对均匀先验分布及非信息先验分布主要基于：总是存在一些先验概率认知可以表达为一个信息先验分布。值得提及的是，研究不能凭空想象，如果确实不存在对某个参数的先验认识，那么该研究就很难进行。

研究是积累性的，贝叶斯分析的主要优点在于它允许使用之前研究分析的结果，这暗示了我们应该寻找先验知识并将这些知识通过信息先验分布整合到研究当中。接下来我们将讨论如何寻找这样的分布以及它们的影响如何。

有关信息先验的一个特别的案例就是当利用贝叶斯推断来复制之前的研究，或者在之前研究基础上有所延伸的时候，我们通常会直接借用之前所研究的后验分布来作为新研究的先验分布，如第 3 章中讲解贝叶斯定理举的例子，我们先有一个含有 1 个观测的样本，然后又收集了一个包含 10 个观测的新样本。例子中，我们将第一次分析所得后验分布作为第二次研究的先验分布，最终所得后验分布与基于 11 个观测的单次分析所得结果相同。

最有用的信息先验是赋予一些参数值零概率，这意味

着这些参数不可能取某些值。例如,根据定义,贝塔分布会把小于0或者大于1的值设定为0,因为参数π只能取两个极端值之间的数值。如果我们设定一些可能的参数值概率为零,就可能出现问题。例如,我们允许回归系数只能取正值,从而对所有负值均赋予零概率。这样做可能出现的问题是,如果我们很肯定参数不能为负,那么,不论经验证据多倾向于负值都于事无补。对于那些先验概率为零的参数值,其后验概率总为零,不论我们可以从数据中得到什么信息。换句话说,由于最初的感觉太强烈,因此不论经验证据多有力,都不会改变我们的想法。

若参数的先验概率为零,那么后验概率也为零,这与贝叶斯定理的性质有关。根据定理,分子包含特定参数值的数据概率与相应参数值的先验概率的乘积。如果特定先验概率等于零,那么分子的整个乘积就为零,即,后验分布等于零。

因此,在给参数值分配零先验概率时必须要谨慎。但是只要先验概率不为零,即便小,我们也有信心将正确的信息记入后验分布。

问题在于我们应该选择非信息先验分布还是信息先验分布,只有在该选择会显著影响后验分布结果时才比较重要。在本章节后面,我们可以看到有关先验分布的影响,数据量越大,先验分布的影响越小,除非我们赋予了某些参数值的子集极小的先验概率。

第2节 | 寻找先验分布

如第3章第一个例子，已知存在三个真实总体，我们可以令每个总体被选中为数据来源的概率都等于1/3。对第一个观测进行贝叶斯分析后，我们又收集了一个包含10个观测的样本。这种情况下，由第一个分析得出后验分布就会被用作第二个分析的先验分布。因而，对于第二个分析，其先验分布就很容易确定。

然而指定先验分布少有如此简单，接下来我们就介绍一下如何寻找先验分布。有关先验分布一个重要的考虑是先验分布应该可以准确地表达参数的先验信息；同时应该容易操作。这两个要求可能有些相悖。对于离散型参数，这并不是一个问题，因为其仅取一些值。相关后验分布的计算，我们可以参考表3.1，且这样的计算也很容易通过计算机编程实现。

当参数被视为连续变量时就可能遇到困难，就如总体百分比或均值的情况。对于这样的参数，先验分布可以绘制为一个连续型曲线，贝叶斯定理可以利用该曲线的数学

函数生成后验分布。从数学上讲,某些函数比其他函数更容易操作,若条件允许,则我们自然会更倾向于这些函数。这些特定先验分布被称为共轭先验分布(conjugate prior distribution)。其实在第 4 章及第 5 章,我们已经有所涉及,尽管它们并没有以此名称命名。

为描述有关共轭先验分布的想法,我们先简单地回顾一下对总体比例 π 的贝叶斯分析。对于一个给定参数值及一个包含 n 个观测和 x 次成功事件的二项分布数据,数据概率可以用二项分布函数表达:

$$f(x \mid \pi) = \binom{n}{x} \pi^{x}(1-\pi)^{n-x} \qquad [6.1]$$

由于数据已知,π 即该式中的唯一未知数。类似的例子可见等式 4.7。其中,贝叶斯定理中的分子,其包含了该等式和先验分布 $f(\pi)$的乘积。

现在问题变成用什么数学函数可以令数据概率(等式 6.1)和先验分布的乘积容易找到。数学为我们提供了大量可能的函数,比如多项式、三角函数以及指数函数等。但是这里我们用的是:

$$f(x) = C'\pi^{a-1}(1-\pi)^{b-1} \quad a>0,\ b>0$$
$$0<\pi<1 \qquad [6.2]$$

其中 C'为通过 a、b 计算所得的常数。相关函数例子请见等式 4.2。指数项中的“-1”基于实际操作考虑,这里我们

大可不必关注。

等式 6.1 和等式 6.2 中乘积项 $f(x|\pi)f(\pi)$容易得出，因我们只需将二项式系数与 C'相乘，然后分别加上 π 的指数及 $1-\pi$ 的指数。基于此操作，所得后验分布包含了有关 π 及 $1-\pi$ 的指数项。相关范例的后验分布可见等式 4.9。因此，π 的先验及后验分布均服从贝塔分布。

如前所提及，当我们将贝塔分布作为总体比例 π 的先验分布，且数据满足二项分布假设时，那么所得 π 的后验分布也服从贝塔分布。因此，当数据为二项分布时，贝塔分布被称为 π 的共轭先验分布。另一个涉及有关共轭先验分布的例子是在分析正态分布的均值时。在第 5 章中，我们知道如果均值的先验分布和所收集的样本数据同时服从正态分布且总体方差已知的情况下，均值的后验分布也服从正态分布。因此，当数据为正态，正态分布也被称为总体均值的共轭先验分布。

由于共轭先验分布可以简化寻找后验分布的方式（所需计算很少，且根据贝叶斯定理可以直接得到后验分布），因此在其存在的情况下，我们会尽量使用共轭先验分布。

但是当共轭先验分布存在的情况下也无法回答有关先验分布设定的问题。例如，对于贝塔分布，我们需要指定参数 a 和 b；对于正态分布，我们需要知道分布的均值和方差。这些常数决定了分布形态，而常数的确定必须经由相关专家认可。

当对总体比例的了解非常有限时，确定 a、b 值似乎会有些困难，但是即便有限也可以生成先验分布。如果某个专家说："基于我的知识积累，该比例差不多在 0.3 和 0.7 之间。"那么，我们就可以说先验分布的均值等于 0.5，标准差等于 0.1。因为从该均匀减去或加上 2 个标准差就是 0.3 和 0.7，然而有时，加上或减去 2 个标准差可以取任意概率。根据等式 4.6，a 和 b 等于 12，那么先验分布就可以确定了。对于各种有助于找到先验分布的有趣而详实的讨论请见瑞法（Raiffa，1968）。

第3节 | 先验的主观性质

一个先验分布可以表达为在数据已知前分析者对总体参数的想法。有关参数的信息通常非常有限，且因人而异。从而，先验分布是主观的，这意味着尽管问题相同，不同人生成的先验分布也可能不同。

不同的先验分布本身不会有什么问题，而当不同的先验分布运用相同的数据时，产生不同的后验分布就是一个问题了。我们怎么可以容许对同样数据得出的统计推断不同呢？

不单单是贝叶斯推断存在主观选择会影响分析的情况。对于经典零假设检验也存在，即我们拒绝或者无法拒绝假设。当数据相同且检验统计量的数值也相同时，一个分析师可能得到结论：该零假设应该被拒绝；而另一个分析师可能得出结论：该零假设无法被拒绝。这时，到底该不该拒绝完全取决于检验统计量的拒绝区间有多大，其中，拒绝区间由显著性水平的大小决定。

由于假设是否被拒绝主要取决于显著性水平的大小，

而显著性水平的选择通常是主观的，从而一个零假设拒绝与否也存在主观性。我们可以争论说，5%显著性水平是所有社会学家检验零假设的通用标准，而同时，我们不能否认拒绝与否完全取决于我们如何选择显著性水平。因此，一种可能的情况是，一个零假设被一位研究者拒绝而不一定被另一位研究者也拒绝。

贝叶斯推断的主观性有些不同。它反映了一个非常人性化的事实，即如果两个人对一件事持不同观点，那么，相同且有限的经验事实可以对不同观点进行不同的修正。这就像一名民主党党员和一名共和党党员面对同一个失业图表时，民主党党员说这张图表明我们需要更多的政府干预及新的公共项目，而共和党党员认为私营企业在这方面做得很好，它们有能力处理好这个问题。类似这样诚恳的争论比比皆是，即使经验事实对双方来讲是一样的。而之所以会有这种区别，原因在于它们所基于的先验信念不同。

由于所受训练与经验不同，对于同一问题，不同社会科学家所用的先验分布也可能不同。尤其当先验分布信息丰富而数据的新信息有限时，后验分布必然不同。然而先验分布将分析的主观性对所有人都开放。这在某种程度上表明，尽管分析者在设定先验分布的时候是基于个人观点，所用先验分布有偏差，但是由于这些信息对所有人都开放，从而其主观性程度实际上减小了。

有关先验分布并无对错,我们所做的就是试图表达有关概率的有限先验知识。分析者所用先验分布是对个人观点的表达,从而没有理由认为为什么每个研究者或者研究报告的听众不能用自己的先验分布。由于基于同样的数据,从而所得的后验分布间的差异并不会如先验分布间那么大。

要降低先验分布的主观程度,我们可以使用非信息化先验分布,比如,均匀分布。这仍然代表特定分布的选择,但是其包含的个人先验观点最少。尽管该分布在大多数情况下都是一个不错的选择,但是当我们对先验分布有一定的认识和足够的信息时,还是应该勇于尝试自己的先验分布。在后验分布主要由先验分布决定的情况下,这种尝试尤为重要。

第4节｜先验的影响

最后一个可能也是最重要的问题就是，先验分布究竟有多重要？它们在概念上不可缺少，且可以对贝叶斯分析提供基础，但是在决定后验分布的确切形态时，在很多情况下都没有那么重要。

检验先验分布影响的一种方法是思考之前章节所提到的一些相关方程。例如，由方程4.16可以得知，在研究总体比例时所得后验均值服从贝塔分布，其可表示为$(a+x)/(a+b+n)$。其中，a和b均来自先验分布，而x和n可以从数据中得出。如果a和b相对于x和n较小，那么对于不同的a和b这部分的值不会有很大变化。这时候，先验分布的影响几乎可以忽略不计。

另一种方法是根据下式等号右边的信息重新写出后验分布：

$$\frac{a+x}{a+b+n}=\frac{\frac{a}{a+b}(a+b)+\frac{x}{n}(n)}{a+b+n}$$

$$=\frac{\text{先验估计}\cdot\text{先验权重}+\text{数据估计}\cdot\text{数据权重}}{\text{权重和}}$$

[6.3]

总体比例先验分布的估计等于 $a/(a+b)$，由数据本身得到的估计为 x/n。在方程 6.3 右边为两个估计的加权平均。先验估计的权重为 $a+b$，由数据得来的估计权重为 n。因 $a/(a+b)$ 和 x/n 都为比例，从而其取值为 0 到 1。且对于等式 6.3，$a+b$ 与 n 的相对大小才是最重要的。值得提及的是，即使信息先验较好，$a+b$ 的值也一般也不会超过 20 或 30，若样本量大小与 $a+b$ 类似，那么先验分布的影响就非常重要。但是若我们所处理数据样本量很大，例如，包含 1 500 个观测的全国性调研，那么在该表达中，数据为主导。此时，先验分布对实质结果的影响几乎为零。

若考虑等式 4.16 的后验方差，可得到相同的结果。若后验均值的取值范围很宽，方差的分子基本为常数，此时影响方差的是则分母的大小。对于方差，其分母等于 $n+a+b+1$，其中 $a+b$ 为先验分布的影响，n 为数据的影响。若 $a+b$ 相对于 n 较小，那么先验分布对后验分布几乎没有影响。值得注意的是，我们所讨论的是 $a+b$ 的和，而不是单独的 a 和 b。如果 n 等于 1 500，$a+b$ 等于 30，则先验分布的影响几乎可以忽略，且对于后验方差的分母，不论 a 等于 1，b 等于 29 还是任意满足 $a+b=30$ 的 a、b 取值，都没有什么区别。

对于其他等式，我们也可以得到类似的结论。例如，分析由正态分布得到的均值时，后验分布的均值等于先验均值和数据均值的加权平均。且其权重为相应方差的倒

数。由于样本量在样本均值方差的分母中，因此对于大样本，先验分布对后验分布仅有很小的影响或者没有影响。

这些结果就是稳定估计(stable estimation)的原则。该原则表明了对于大样本，即使存在信息非常丰富的先验分布，其对后验分布的影响也几乎可以忽略。原因在于如果样本很大，除了一些位于狭窄取值范围内的参数值，数据的概率都趋近于0。例如，由方程4.7得出了，在含有1 830个观测的样本中，有420个天主教徒和1 410个非天主教徒的概率。概率的二项系数为一常数，且被贝叶斯定理所吸收，数据概率的重要部分包含在了$\pi^{420}(1-\pi)^{1\,410}$表达式中，在统计上称为似然函数(likelihood function)。每个π值都会与贝叶斯定理分子中相同π值下所对应的先验分布值相乘。但是当π小于0.2或者大于0.26时，该值几乎等于0，且位于这两个值构成区间内的分布非常陡峭。这表明对于$\pi<0.2$或$\pi>0.26$，不论同区间的先验分布取何值，后验分布都基本为零。从而，先验分布的影响只有在区间0.20—0.26才会体现，且大多先验在这样狭窄的区间内均服从均匀分布。因此，后验分布的形状基本完全由似然函数决定，且大多区间内后验分布所包含的参数值均是数据所倾向的。

稳定估计占了贝叶斯统计和经典统计之间的大多数对应数值。这就解释了为什么贝叶斯概率区间与经典置信区间经常在数值上重合，从而我们可以将置信区间解释

为贝叶斯区间。这还解释了为什么在数据中存在强证据时,不同的先验分布仍然会得到相同的后验分布。

反之亦然。当数据的信息有限时,即,数据的样本量小或者方差大,或者两者都有,此时,先验分布不同,所得的后验分布也不同。而后验分布并没有像先验分布那样不同,因为先验分布可以被同样的数据修改。但是问题在于,数据中的信息也不一定足以令不同的先验分布汇集成一个单一后验分布。

第7章

贝叶斯的难点

贝叶斯定理通过结合先验概率分布和条件数据概率计算参数的后验分布。其中,可能遇到的困难包括寻找先验分布、寻找数据概率以及操纵贝叶斯定理以找到合适的后验分布。这一章我们会仔细讲解贝叶斯统计中这三种常见的问题。

第 1 节 | 先验

寻找一个合适的先验分布可能很难。因为只要我们对一个参数有一些模糊的先验认识时，就需要将这些认知转化为一个概率分布。在第 6 章我们已经做了相关讨论，该过程确实很难。

特别是在多元分析中寻找一个信息先验更难。例如，多元回归分析包括截距、每个变量的回归系数以及残差项的方差。有关参数的一个较合适的分布由一个包含所有我们所知道的参数以及参数之间的相关性信息的多元分布组成。即便在只有很少一些参数的时候，这样的联合分布都很难找到；当我们有很多参数时，这样的联合分布几乎不可能找到。

对于多元分析，大多数情况下我们会采用均匀、非信息化且独立的先验分布。这时，我们就可避免表达参数间可能的相关性而带来的困难，从而联合后验分布则主要包含了有关从数据中得到的参数信息。

第 2 节 | 数据概率

第 5 章有关贝叶斯方法的范例主要基于假设:数据是正态分布的。该假设令我们可以在贝叶斯定理中将等式用正态分布取代得到数据概率,然后通过数学语言推导出后验分布。而在总体比例的例子中,我们基于假设:数据服从二元分布,并在贝叶斯定理中运用有关二元分布的等式以求得后验分布。

如果在贝叶斯定理中数据不服从用来生成数据概率的特定分布怎么办?若真实分布偏离所假设的数据分布,在多大程度上会影响所得后验分布?如果采用其他数据概率,我们自由选择这些概率的程度又是如何?我们知道,检验偏离假设数据概率的敏感度,称为稳健性问题。该问题是贝叶斯统计学家与经典统计学家都非常关注的。我们通常用偏度和峰度来研究曲线偏离正态的程度,正态分布则是对称的,且峰度为 0(峰度反映的是某一分布的尾部有多厚)。

贝叶斯处理偏离正态的方法包含用一个比正态分布

更普遍更常见的数据概率。除了中心值和分散程度，该分布通常有包含偏度和峰度参数。根据参数的先验分布就可能导出其后验分布。该过程相对这里讨论的方法更加复杂，感兴趣的读者可以参考博克斯和蒂奥（Box & Tiao，1973），在书中他们对一些偏离正态的模型有所讨论。

关于贝叶斯稳健性的深入讨论可参考伯杰（Berger，1980）。

第3节 | 计算

通过共轭先验分布,贝叶斯定理可以轻松生成后验分布。在这种情况下,定理本身并未得以体现。而当运用更加复杂的先验分布时,我们可能需要回到原理一,运用贝叶斯定理导出后验分布。

如第3章所举例子,当参数为离散型时,我们需要计算表3.1中所有参数值的后验概率。而当参数为连续型时,我们可以用积分代替数值方法来计算参数的后验分布。在计算机技术飞速发展的今天,这并不是一件困难的事,因此如果信息足够,我们不要害怕尝试特殊的先验分布,尽管这些先验分布需要通过复杂的数学计算才可得到参数后验分布。



第8章 贝叶斯的优势

经典统计方法主要关注的是如何通过样本推断总体，尤其是假设检验。在这章我会就一些贝叶斯推断的优势进行讨论，并解释为什么我们运用贝叶斯统计推断，不论在研究或是在训练未来的社会科学家上。

有关运用贝叶斯统计方法的原因存在于两个层面，特定层面和一般层面。从特定层面上说，因为有些问题不能通过经典方法解决，但是用贝叶斯方法就完全没问题。如第 5 章所指出的，比较方差不同的两个总体均值。另一个例子就是下面即将介绍的，被称为停止规则问题。它常用来处理研究许多抽样方案下的总体比例。从综合层面上说，贝叶斯推断是专门为学习研究过程设计的，其始于对参数最初的不确定性，然后基于数据中新的信息对该不确定性进行更新。而对于使用双重否定的经典统计方法就不可能存在这种不确定的情况，即假定研究假设为真，然后检验数据与该假定是否一致。

第1节 | 特定层面原因

不同均值，不同方差

一个统计上没有确切的传统的解决方案的问题就是当总体方差不同时，如何检验总体均值没有差异的零假设。尽管答案在数学上非常复杂，但是用近似方法就会简单很多(如第5章所示)，贝叶斯统计即提供了这样一种研究两均值差异的方法。

停止规则

就获得研究结论而言，我们的意图可以发挥多大作用，数据又可以起到多大用处呢？经典统计认为通过数据本身可以反映问题，这就意味着对他们来说先验分布没有存在的必要。

让我们检验一下这种情况。对于零假设——总体中

有 30%的人倾向于增加军事花费。在一个包含 $n=10$ 个人的随机样本中，我们发现 $x=1$ 个人有这种倾向。根据经典统计推断的发展趋势，这里我们将报告单尾 p 值并让读者自己来决定拒绝与否，而并非选择一个显著性水平来判定零假设是否被拒绝。p 值为可观测数据的概率及其他更极端的数据可能发生的概率，也就是零假设可能被拒绝的最小显著水平。

如前所提及，观测样本中有一个人倾向于增加军事花费，一个更极端的情况可能是没有一个人有这种倾向。通过 $n=10$、$\pi=0.3$ 的二项分布可以发现，0 个人或 1 个人有这种倾向的概率为 0.15，按照大多数标准，该概率值并不足以小到结果显著，即，无法拒绝零假设。因此，根据经典统计解释，我们可以知道，如果从该总体中抽出足够多的含有 10 个值的样本，假设零假设为真，其中有 15%的样本里有 0 个或者 1 个人支持增加军费开支。

此时情况变得更加复杂，因为研究者在告诉我们："事实上，我想要一个样本，其中有一个人倾向于增加军事话费，然而在找到这样一个人之前，我必须抽至少 10 人。"换句话说，即使我们处理的是同一个数据，且研究者希望存在 $x=1$ 个人有这种倾向的情况是固定的，事实上却是样本量 10 存在随机性。也就是说，在另一个样本中，为了找到 1 个人有这种倾向，样本量可能得达到 7 或 11 或其他数值。观测到的数据包括 $n=10$ 个人，更极端的数据可能包

含11或12或者更多的人。每一个可能的n值的概率都可以通过负二项分布找到，且这些概率的之和等于p值，这里即0.05，在单尾检验中边际显著。该概率的经典统计学解释，即假设零假设为真，若我们抽取大量样本，且每个样本均包含有一个倾向增加军事花费的人，那么，有5%的样本的样本量可能大于或者等于10。

对于第一个例子，p值等于0.15；对于第二个例子，相应p值等于0.05。因此，即便在两个例子中用的都是同一个数据，一个p值是另一个的三倍。在观测数据中，其中一个人倾向增加军事花费而其他9个不是。两个例子的唯一区别就在于，第一个例子里，我们希望取10个观测，并对其中倾向军事花费的人计数；而在第二个例子中，我们希望一直抽样直到样本中有一个人存在这种倾向。

更加复杂情况是当研究者简单地收集了一些数据，其最终样本量为10，且有一个人倾向增加军事花费，但是他们却并没有计划收集更多含有10个值的样本，或者持续抽样直到抽到一个有该倾向的人。该数据就相当于抽样过程突然停止所得到的数据。在这种情况下，用经典统计方法就无法进行分析。没有关于数据收集目的的信息就无法使用形式化模型(formal model)，从而统计分析也就不可能。

但是贝叶斯推断并不在意数据的收集是否在得到10个观测后就停止了，不管是因为发现了一个倾向军事花费

的人还是其他，重要的是数据本身。对于数据，其中有一个人有这种倾向而其他九个人没有，根据该信息，我们可以将其用数学语言表达成 $\pi^{1}(1-\pi)^{9}$，或者更一般的表达为 $\pi^{x}(1-\pi)^{n-x}$。令该似然函数与先验分布结合，若先验服从贝塔分布，那么由第 4 章相关内容可知后验分布也服从贝塔分布。

第 2 节 | 一般层面原因

贝叶斯统计推断方法优于经典统计推断的一个原因在于贝叶斯统计的累积特征。我们可以在先验分布中表达对参数的先验认识这一事实表明，分析不必每次都重新开始。在很多时候，当确有先验信息时，信息先验的使用就成为可能。当有了信息先验分布，所得后验分布的峰更高方差更小。这意味着我们可以更精确得估计未知参数，与非信息化先验相比，可得到更小的参数贝叶斯概率区间。

当数据持续收集时就可能存在先验信息。如芝加哥大学的美国综合社会调查和密歇根大学的美国大选研究都进行了多年，其均可为之后的研究分析提供相关信息。为了适应时变趋势研究，每次调查都存在重复问题，尽管如此，先前丰富的研究信息均可以用作我们的先验分布。

贝叶斯统计的另一个优点是其适合教学。大多社会科学家在统计方法上的训练都很有限。一个典型的社会科学研究生项目包含两或三门统计课，而大多本科生项目

更少。材料覆盖了大部分内容，但是进行统计推断的机会很少。尤其对于那些数学天赋有限的研究者，他们对类似显著性水平、检验功效、抽样分布以及置信区间等概念均难以理解。

引入不确定性这个一般性的概念，贝叶斯推断更加客观，其结论更容易理解。其实，不仅仅是学生，有些统计教科书的作者对诸如置信区间之类的概念，理解都存在困难。而在学习贝叶斯概率区间时，类似的困难却不存在。

这点在没有接受过正统统计学训练却又要用统计结果的人中尤为突出。如，在法庭上，当一个统计顾问被聘为专家时，陪审员、律师及法官均需要通过他来了解统计学家的研究结果。凯(Kaye, 1982)讨论了由同一数据得出的五个可能的统计分析指标：p 值、零假设的拒绝、预期区间、似然函数和贝叶斯方法。他绕过了有关先验分布的问题，且并不推荐使用贝叶斯方法。在类似情况下，我个人的经验表明贝叶斯方法在教学与交流上有显著的优势。

最后，贝叶斯推断更加直观的另一个原因在于，相对于经典推断，它更加接近研究过程本身。研究问题通常始于对一个或多个参数的不确定性，然后我们收集数据来增加对参数的了解，基于新的信息，最初的不确定性就减少了。

对研究的描述同时也是对贝叶斯统计推断的描述。其始于参数的先验分布，基于给定的各种参数值计算数据

概率,然后根据贝叶斯定理将所得数据概率与先验分布结合,结果就等于参数的后验分布。贝叶斯定理利用概率的概念来表达我们对真实数值的不确定性,即,未知参数。而经典推断则通过观测数据计算概率,观测数据并不存在不确定性。它们就是大家所了解和所看到的。

在经典统计中,当零假设被拒绝时,我们有多少人这样思考过,“尽管不能非常确定,但这很可能意味着零假设是假的”。不论是与否,我们常常直接运用贝叶斯概率解释结果。然而经典统计并不允许以上陈述。零假设可真可假,只是真实情况如何我们不知道而已。然而由于概率被定义为相对频率,因此零假设并不总为真。

通过考虑这种思维方式所带来的后果和贝叶斯方法的整个分析过程,我们应该诚实得对待经典统计推断所得别的结果,而不是直接用贝叶斯概率对其进行解释;我们应该承认参数的不确定性,并通过贝叶斯统计推断过程来处理该不确定性。

参考文献

BARNETT, V. (1982) Comparative Statistical Inference. New York: John Wiley.

BAYES, T. (1763) "Essay towards solving a problem in the doctrine of chances." Philosophical Transactions. Royal Society, London 53: 370—418. (Reprinted in Biometrika, 1958, 45: 293—315.)

BERGER, J. (1980) Statistical Decision Theory: Foundations, Concepts and Methods. New York: Springer-Verlag.

BERNARDO, J. M., M. H. DeGROOT, D. V. LINDLEY, and A.F.M. SMITH [eds.] (1980) Bayesian Statistics. Proceedings of the First International Meeting, Valencia, Spain. Valencia, Spain: University Press.

BOX, G.E.P. and G.C. TIAO (1973) Bayesian Inference in Statistical Analysis. Reading, MA: Addison-Wesley.

CAMPBELL, A. and P. E. CONVERSE (1980) Quality of American Life, 1978. Conducted by the Institute for Social Research, University of Michigan, ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. (machine readable data file)

DeFINNETTI, B. (1982) "Probability and my life," in J. Gani (ed.) The Making of Statisticians. New York: Springer-Verlag.

EDWARDS, W., H. LINDMAN, and L. J. SAVAGE (1963) "Bayesian statistical inference for psychological research." Psychological Review 70 (May): 193—242.

ELDER, G. H., Jr. (1969) "Appearance and education in marriage mobility." American Sociological Review 34: 524.

HENKEL, R. E. (1976) Tests of Significance. Beverly Hills, CA: Sage.

IVERSEN, G. R. and H. NORPOTH (1976) Analysis of Variance. Beverly Hills, CA: Sage.

JEFFREY, R. C. (1983) The Logic of Decision. Chicago: University of Chicago Press.

JEFFREYS, H. (1961) Theory of Probability. Oxford: Clarendon Press.

KAYE, D. (1982) "Statistical evidence of discrimination." Journal of the American Statistical Association 77 (December): 773—783.

KYBURG, H. E., Jr. and H. E. SMOKLER [eds.] (1980) Studies in Subjective Probability. Huntington, NY: R. E. Kreiger.

LINDLEY, D. V. (1965) Introduction to Probability and Statistics from a Bayesian Viewpoint (2 vols.). Cambridge: Cambridge University Press.

PHILLIPS, L. D. (1974) Bayesian Statistics for Social Scientists. New York: Crowell.

RAIFFA, H. (1968) Decision Analysis. Reading, MA: Addison-Wesley.

RAMSEY, F. P. (1926) "Truth and probability." in The Foundation of Mathematics and Other Logical Essays. London: Routledge & Kegan Paul. (Reprinted in Kyberg and Smokler [1964].)

ROSENKRANTZ, R. D. (1977) Inference, Method and Decision. Towards a Bayesian Philosophy of Science. Dordrecht, Holland: D. Reidel.

SAVAGE, L. J. (1954) The Foundations of Statistics. New York: John Wiley.

SCHMITT, S. A. (1969) Measuring Uncertainty. An Elementary Introduction to Bayesian Statistics. Reading, MA: Addison-Wesley.

VENN, J. (1886) The Logic of Chance. London: MacMillan. (Reprinted by Chelsea, New York, 1963.)

译名对照表

Bayesian statistical inference	贝叶斯统计推断
beta distribution	贝塔分布
binomial distribution	二项分布
bivariate normal distribution	二元正态分布
confidence interval	置信区间
conjugate prior distribution	共轭先验分布
contingency table	列联表
formal model	形式化模型
formal theory	形式理论
Gamma function	伽马函数
harmonic mean	调和平均数
informative prior	信息先验
likelihood function	似然函数
stable estimation	稳定估计
tail probabilities	尾概率
two-sided test	双向检验
uniform distribution	均匀分布

格致方法・定量研究系列

1. 社会统计的数学基础
2. 理解回归假设
3. 虚拟变量回归
4. 多元回归中的交互作用
5. 回归诊断简介
6. 现代稳健回归方法
7. 固定效应回归模型
8. 用面板数据做因果分析
9. 多层次模型
10. 分位数回归模型
11. 空间回归模型
12. 删截、选择性样本及截断数据的回归模型
13. 应用 logistic 回归分析（第二版）
14. logit 与 probit：次序模型和多类别模型
15. 定序因变量的 logistic 回归模型
16. 对数线性模型
17. 流动表分析
18. 关联模型
19. 中介作用分析
20. 因子分析：统计方法与应用问题
21. 非递归因果模型
22. 评估不平等
23. 分析复杂调查数据（第二版）
24. 分析重复调查数据
25. 世代分析（第二版）
26. 纵贯研究（第二版）
27. 多元时间序列模型
28. 潜变量增长曲线模型
29. 缺失数据
30. 社会网络分析（第二版）
31. 广义线性模型导论
32. 基于行动者的模型
33. 基于布尔代数的比较法导论
34. 微分方程：一种建模方法
35. 模糊集合理论在社会科学中的应用
36. 图解代数：用系统方法进行数学建模
37. 项目功能差异（第二版）
38. Logistic 回归入门
39. 解释概率模型：Logit、Probit 以及其他广义线性模型
40. 抽样调查方法简介
41. 计算机辅助访问
42. 协方差结构模型：LISREL 导论
43. 非参数回归：平滑散点图
44. 广义线性模型：一种统一的方法
45. Logistic 回归中的交互效应
46. 应用回归导论
47. 档案数据处理：生活经历研究
48. 创新扩散模型
49. 数据分析概论
50. 最大似然估计法：逻辑与实践
51. 指数随机图模型导论
52. 对数线性模型的关联图和多重图
53. 非递归模型：内生性、互反关系与反馈环路
54. 潜类别尺度分析
55. 合并时间序列分析
56. 自助法：一种统计推断的非参数估计法
57. 评分加总量表构建导论
58. 分析制图与地理数据库
59. 应用人口学概论：数据来源与估计技术
60. 多元广义线性模型
61. 时间序列分析：回归技术（第二版）
62. 事件史和生存分析（第二版）
63. 样条回归模型
64. 定序题项回答理论：莫坎量表分析
65. LISREL 方法：多元回归中的交互作用
66. 蒙特卡罗模拟
67. 潜类别分析
68. 内容分析法导论（第二版）
69. 贝叶斯统计推断



贝叶斯方法是一种计算假设概率的方法。该方法将未知模型或变量看成已知分布的随机变量，基于先验和后验概率对未知参数进行推断。作为进行决策的重要工具，贝叶斯方法被广泛应用于解决医学、金融、市场预测、产品检测等一系列不确定的问题。

主要特点

- 与经典统计方法比对，有助于读者掌握贝叶斯方法基本思想
- 案例丰富，语言通俗，方便读者理解贝叶斯理论及实际应用
- 适合具有经典统计方法基础的高年级本科生、研究生


上架建议：社会研究方
ISBN 978-7-5432-287
9 787543 228764

您可以通过如下方式联系到我们：
邮箱：hibooks@hibooks.cn


微信


天猫

定价：32.00元
易文网：www.ewen.co
格致网：www.hibooks.cn